The ICON Earth System Model Version 1.0

J.H. Jungclaus[1], S.J. Lorenz[1], H. Schmidt[1], V. Brovkin[1,6], N. Brüggemann[1,2], F. Chegini[1], P. De-Vrese[1], V. Gayler[1], M.A. Giorgetta[1], O. Gutjahr[1,2], H. Haak[1], S. Hagemann[5], M. Hanke[3], T. Ilyina[1], P. Korn[1], J. Kröger[1], L. Linardakis[1], C. Mehlmann[1], U. Mikolajewicz[1], W.A. Müller[1], J.E.M.S Nabel[1,*], D. Notz[2,1], H. Pohlmann[1,4], D.A. Putrasahan[1], T. Raddatz[1], L. Ramme[1,7], R. Redler[1], C.H. Reick[1], T. Riddick[1], T. Sam[1], R. Schneck[1], R. Schnur[1], M. Schupfner[3], J.-S. von Storch[1,6], F. Wachsmann[3], K.-H. Wieners[1], F. Ziemen[1,3], B. Stevens[1], J. Marotzke[1,6], and M. Claussen[1,6]

[1]Max-Planck-Institute for Meteorology, Hamburg, Germany
[2]Institute of Oceanography, Universität Hamburg, Hamburg, Germany
[3]Deutsches Klimarechenzentrum, Hamburg, Germany
[4]Deutscher Wetterdienst, Hamburg, Germany
[5]Helmholtz Zentrum Hereon, Geesthacht, Germany
[6]Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Germany
[7]International Max Planck Research School on Earth System Modelling, Hamburg, Germany
[*]now at: Max Planck Institute for Biogeochemistry, Jena, Germany

Key Points:

- This work documents ICON-ESM 1.0, the first version of a coupled model based on the ICON framework
- Performance of ICON-ESM is assessed by means of CMIP6 DECK experiments at standard CMIP-type resolution
- ICON-ESM features good performance in the stability of TOA radiation balance and reproduces the observed temperature evolution. Biases in clouds, winds, sea-ice, and ocean properties are somewhat larger than in MPI-ESM at similar resolution.

Corresponding author: Johann Jungclaus, johann.jungclaus@mpimet.mpg.de

Abstract

This work documents the ICON-Earth System Model (ICON-ESM V1.0), the first coupled model based on the ICON (ICOsahedral Non-hydrostatic) framework with its unstructured, isosahedral grid concept. The ICON-A atmosphere uses a nonhydrostatic dynamical core and the ocean model ICON-O builds on the same ICON infrastructure, but applies the Boussinesq and hydrostatic approximation. The oceanic carbon cycle and biogeochemistry is represented by the HAMOCC6 module and the terrestrial biogeophysical and biogeochemical process are integrated in the new JSBACH4 module. We describe the tuning and spin-up of a base-line version at a resolution typical for models participating in the Coupled Model Intercomparison Project (CMIP). The performance of ICON-ESM is assessed by means of a set of standard CMIP6 simulations. Achievements are well-balanced top-of-atmosphere radiation, stable key climate quantities in the control simulation, and a good representation of the historical surface temperature evolution. The model has overall biases, which are comparable to those of other CMIP models, but ICON-ESM performs less well than its predecessor, the MPI-ESM. Problematic biases are diagnosed in ICON-ESM in the vertical cloud distribution and the mean zonal wind field. In the ocean, sub-surface temperature and salinity biases are of concern as is a too strong seasonal cycle of the sea-ice cover in both hemispheres. ICON-ESM V1.0 serves as a basis for further developments that will take advantage of ICON-specific properties such as spatially varying resolution, and coupled configurations at very high resolution.

Plain Language Summary

ICON-ESM is a completely new coupled climate and earth system model that applies novel design principles and numerical techniques. This article describes how the component models for atmosphere, land, and ocean are coupled together and how we achieve a stable climate by setting certain tuning parameters and performing sensitivity experiments. We evaluate the performance of our new model by running a set of experiments under pre-industrial and historical climate conditions as well as a set of idealized greenhouse-gas-increase experiments. These experiments were designed by the Coupled Model Intercomparison Project (CMIP) and allow us to compare the results to those from other CMIP models and the predecessor of our model, the Max Planck Institute for Meteorology Earth System Model. While we diagnose overall satisfying performance, we find that ICON-ESM features somewhat larger biases in several quantities compared to its predecessor at comparable grid resolution. We emphasize that the present configuration serves as a basis from where future development steps will open up new perspectives in earth system modelling.

1 Introduction

ICON-ESM (V1.0) is the first release of a new Earth System Model that is developed at the Max Planck Institute for Meteorology (MPI-M). It is based on the ICON framework, a joint development of MPI-M, the German Weather Service (Deutscher Wetterdienst, DWD), the Karlsruhe Institute for Technology, and other partner institutions in Germany and Switzerland. ICON-ESM combines the ocean ICON-O (Korn, 2017) and atmosphere ICON-A (Giorgetta et al., 2018) components of the ICON modelling system together with ICON-Land, including JSBACH 4, a complete re-write of the land model JSBACH 3 (Reick et al., 2021, 2013), and the ocean biogeochemistry module HAMOCC6 (Ilyina et al., 2013). The ocean and atmosphere are coupled using the newly developed coupling software Yet Another Coupler (YAC; Hanke et al., 2016). ICON stands for ICOsahedral Nonhydrostatic, where the latter is only realized for the atmosphere component.

At MPI-M, ICON-ESM succeeds the well-established Max Planck Institute for Meteorology Earth System Model (MPI-ESM; Mauritsen et al., 2019) with its component models for the atmosphere ECHAM6 (Stevens et al., 2013) and the ocean MPIOM (Jungclaus

et al., 2013), the land model JSBACH 3 (Reick et al., 2013, 2021), and the ocean biogeochemistry module HAMOCC (Ilyina et al., 2013). Together with its predecessors, MPI-ESM1.2 has represented three decades of successful model development (see Mauritsen and Roeckner (2020)). The development of a completely new model system is an answer to the requirement for increasing resolution, the need for conservation for the representation of chemical tracers in the atmosphere, and for excellent scalability at high-performance computers (HPC). While these properties of ICON-ESM will be most beneficial in very high-resolution coupled configurations, we present here, as a first step to introduce ICON-ESM to the scientific community, the physical model at a resolution that can be called "standard" in the context of climate simulations for the ongoing Coupled Model Intercomparison Project (CMIP6, Eyring et al. (2016)). We focus on a set-up with 158 km grid spacing in ICON-A and 40 km in ICON-O. Focusing on experiments in climate mode, i.e. at least century-long simulations with parameterized physics, the set-up described here offers an efficient configuration for simulations of past, present and future climates, and large ensembles. It also forms the basis for higher-resolution versions as well as for configurations using specific properties of the ICON system, for example grid refinement in ICON-O (Logemann et al., 2021) or nesting in ICON-A (Klocke et al., 2017). In this manuscript, we present the first results of ICON-ESM and provide an examination of the model characteristics in a set of experiments following the CMIP6 Diagnosis, Evaluation, and Characterization of Klima (DECK) protocol and include an ensemble of five CMIP6 "historical" simulations (Eyring et al., 2016). We compare and evaluate the simulations with observations and reanalysis data as well as other CMIP6 models and MPI-ESM.

Typically, model tuning happens initially at the component model level (Giorgetta et al. (2018), Korn P. et al., "ICON-O: The ocean component of the ICON Earth System Model - global simulation characteristics and local telescoping capability"). Coupled together, the completely new ICON components for ocean, sea-ice, land and atmosphere repeatedly revealed unexpected behavior that required detailed investigations and major tuning efforts, which we partly describe in this manuscript (section 3). In the following, we provide information on the general circulation models for atmosphere and ocean, the land and sea-ice components, the ocean biogeochemistry module, and the coupler. Then we describe the spin-up and tuning of the coupled system that has led to the pre-industrial control simulation (piControl) under constant forcing agents. Evaluation in comparison with observations and reanalyses data is based on the last decades of a small ensemble of CMIP6 historical simulations and the model's climate sensitivity characteristics are assessed in idealized global warming experiments (i.e., the 1 %CO_2 yr^{-1} increase experiment (1pctCO2) and the experiment with an abrupt four-fold CO_2 concentration (abrupt4xCO2)). We discuss tuning choices in section 5 and end with a summary and conclusion (section 6).

2 Model overview

The ICON model system (Zängl et al., 2015) provides common infrastructure (e.g. grid construction and output handling) and, in part, common numerical operators for the component models. A common feature is the basic grid construction based on unstructured, icosahedral grids. The grids for both the ICON-A and ICON-O model are created by recursively dividing the original twenty triangles of the icosahedron. This is done by bisecting the edges. The vertices at each step are projected at the Earth sphere. The primary cells are triangles, while the dual cells are hexagons, except for the original twelve pentagons of the icosahedron which remain. A detailed description of the process is given in H. Wan et al. (2013) and Giorgetta et al. (2018). The spring dynamics optimization is applied on both grids, as described in Tomita et al. (2001) and Tomita et al. (2002). The grids are symmetrized with respect to the equator by reflecting the northern hemisphere to the south. The equatorial-symmetric grid has been tested with

the ICON-O for shallow water set-ups and showed reduced errors (Korn & Linardakis, 2018). Local asymmetries in grids can be the cause of increased numerical errors (Weller et al., 2009). For the icosahedron, these asymmetries occur most profoundly in the vicinity of the pentagons (Korn & Linardakis, 2018). It is desirable to keep these "hot" spots away from areas where large velocities may occur, for example due to the orography. Therefore the grid was rotated 37^o eastwards, to avoid placing a pentagon over the Himalaya region. In the set-up presented here, the resolution for the ICON-A grid is 158 km, measured as the square root of the average triangle area, with a total of 20480 triangles (the R2B4 grid in Table 1 of Giorgetta et al. (2018)). The ICON-O grid has an average resolution of 40 km and 235403 triangles, the land triangles being removed to reduce memory and computing resources. The bathymetry was interpolated from the SRTM3 PLUS dataset (Becker et al., 2009), adjusted to conform with the sea-land mask given by the GLCC 2.0 dataset (Global Land Cover Characterization (GLCC), doi:10.5066/F7GB230D). The ICON-O sea-land mask is then projected to the coarser ICON-A grid, allowing for triangles to be partially ocean.

The numerical schemes of the atmosphere and the ocean share commonalities but feature also significant differences. Identical in both components is the spatial discretization of differential operators such as divergence and curl through mimetic methods (cf. Korn (2017)). This takes advantage of identical grid structures and the staggering of variables. The staggering necessitates reconstructions and interpolations to connect variables that are located at different grid positions to calculate fluxes. This is accomplished in ICON-O by the novel concept of *Hilbert space admissible reconstructions* (for details see Korn (2017), Korn and Linardakis (2018)). In contrast, ICON-A relies on several interpolation methods (see Zängl et al. (2015)).

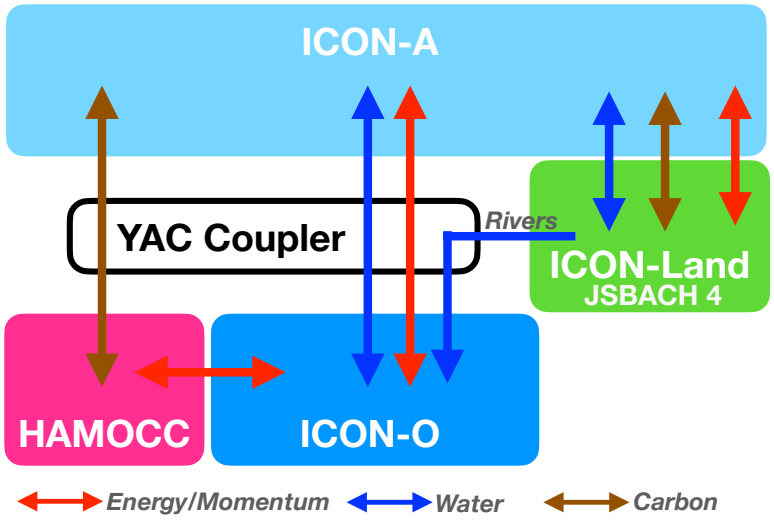

Figure 1. Schematic representation of the model components of the ICON-ESM and the coupling scheme.

2.1 Atmosphere

The atmosphere component of the ICON-ESM is the icosahedral nonhydrostatic atmospheric general circulation model ICON-A. The model version used here is similar to version 1.3.00 described in detail by Giorgetta et al. (2018), which was evaluated by Crueger et al. (2018). Modifications with respect to this earlier version are described be-

low. The dynamical core of the model (Zängl et al., 2015) and the transport scheme are shared with a configuration used for numerical weather prediction (NWP) at the DWD. Other variants of the ICON atmosphere model include the option to interactively couple to the Aerosol and Reactive Trace gases scheme ART (Rieger et al., 2015) and a configuration including the upper atmosphere (UA-ICON, Borchert et al., 2019). The excellent scaling capabilities of ICON have enabled global storm-resolving simulations down to a horizontal grid resolution of about 2.5 km (Stevens et al., 2019).

The model configuration used here differs from the NWP variant in particular with respect to the physics package, which was adopted from the ECHAM6 general circulation model (Stevens et al., 2013) used in the MPI-ESM (Giorgetta et al., 2013; Mauritsen et al., 2019). The physics parameterizations include the PSrad radiation scheme of Pincus and Stevens (2013), a scheme for turbulent vertical diffusion based on a total turbulent energy approach as proposed by Mauritsen et al. (2007), a convection parameterization based originally on the Tiedtke (1989) mass flux scheme, a parameterization for the representation of stratiform clouds including microphysics based on a scheme by Lohmann and Roeckner (1996) and cloud cover diagnosed following Sundqvist et al. (1989), a representation of the effects of gravity waves and blocking from sub-grid scale orography following Lott (1999), and the Hines (1997) parameterization of the effects of non-orographic gravity waves. Adaptations of the original ECHAM parameterization schemes for the use in ICON-A are described by Giorgetta et al. (2018). As in the latter publication, we are using ICON-A here with a horizontal grid resolution of 158 km. In the vertical, the model employs a terrain following hybrid sigma-height grid with 47 layers extending to a model lid at 83 km.

Compared to the ICON-A version 1.3.00 described by Giorgetta et al. (2018) the following modifications have been made. (A) The coupling of the physical processes has been serialized completely using the following sequence: (1) radiative effects by terrestrial longwave and solar shortwave radiation, (2) vertical diffusion with implicitly coupled land surface processes, (3) non-orographic gravity wave drag, (4) subgrid-scale orographic (SSO) effects, (5) cumulus convection, and (6) cloud microphysics. This improved the numerical stability and allowed to increase the model time step from 10 to 15 minutes. Further, the time step for radiation, the only process not computed at every model time step, was shortened from 120 to 90 minutes. (B) The non-orographic gravity wave and SSO effects were re-tuned. Here, new SSO parameters for the statistical description of the unresolved terrain were used, which resolve both issues discussed in section 4.7.1 of Giorgetta et al. (2018), i.e. the error in the azimuthal angle of the unresolved mountains, and the standard deviation of unresolved orographic height, which is now computed with respect to the resolved sloped terrain. Further, a weighting factor for the non-ocean fraction has been introduced to account for the fact that the SSO parameters are computed for the area fraction that is land or lake or glacier, i.e. non-oceanic. And (C) the physical processes were re-tuned for a balanced top-of-atmosphere (TOA) radiation balance and in order to minimize the systematic errors in AMIP simulations.

In step (B), tuning parameters (see Table 1) G, which scales the magnitude of the orographic gravity wave drag, and C_d, which scales the blocking of low-level flow by unresolved orography, were tested with values in the range of 0.01 to 1 with the following goals: The first target was to reduce the systematic error in zonal mean zonal wind in DJF at 60^o N at 10 hPa. The secondary target was then to minimize errors in the zonal mean zonal wind in JJA as well as errors in annual mean pressure at sea level and annual mean zonal wind stress at the ocean surface. This led to new default parameters $G = 0.05$ and $C_d = 0.05$, instead of $G = 0.10$ and $C_d = 0.01$ (Giorgetta et al., 2018). The tuning parameters for the non-orographic gravity wave drag remained as in Giorgetta et al. (2018).

In step (C) a range of tests was conducted with modifications in tuning parameters for fractional cloud cover, entrainment of environmental air in convective plumes,

overshooting mass flux fraction at the top of convection, and cloud microphyiscs. From all tests a configuration with three modifications was chosen, compared to Giorgetta et al. (2018): The entrainment coefficients for deep and shallow convection were set to $\epsilon_{shallow} = \epsilon_{pen} = 0.0003$ m^{-1}, and the convective mass flux fraction across the level of neutral buoyancy at the top of convection was reduced to $c_{mfctop} = 0.1$. This configuration performed best following a similar evaluation as presented in Giorgetta et al. (2018). However, it should be noted that other tested configurations were equally acceptable concerning the radiation balance at the top of the atmosphere, which was the primary tuning goal.

The resulting atmospheric model configuration provided the starting point for the development of the coupled model system in the pre-industrial control experiment, which lead to additional changes of tuning parameters for dynamics as well as physics, as described in Section 3 and reviewed in the discussion section.

2.2 Ocean

ICON-O, the ocean general circulation model that provides the ocean component of ICON-ESM, solves the hydrostatic Boussinesq equations. These dynamical equations are also referred to as the "primitive equations". The state vector consists of horizontal velocity, the oceanic tracers potential temperature and salinity, as well as the surface elevation, due to the free surface boundary condition. The primitive equations are solved on the triangular ICON grid with an Arakawa C-type staggering that places tracers at the circumcenter of a triangular cell and the normal component of the velocity vector at the midpoint of the cells edge. The vertical coordinate-axis is given by the z-coordinate (or geopotential height). The two-dimensional triangles are simply extended by a height-based dimension. This generates three-dimensional prisms. The number of vertical levels depends on the topography and varies from cell to cell $N_z = N_z(K)$. The thickness of the prisms is constant, except for the surface layer, where the sea surface elevation is taken into account. Alternative vertical coordinates such as the z^*-coordinate are available in ICON-O and are described in Sing and Korn, manuscript in preparation "A structure-preserving discretization of ocean models in generalized vertical coordinates". In the work presented here we use z-coordinates as the classical choice with well-understood advantages and disadvantages.

The subgrid scale closure for velocity uses a biharmonic operator based on the vector Laplacian with a viscosity coefficient that scales with the square root of edge length times cell center distance to the third power. Eddy-induced diffusion and eddy-induced advection are parameterized following Redi (Redi, 1982) and Gent-McWilliams (GM) (P. Gent & McWilliams, 1990), respectively. We employ the variational approach of Griffies et al. (1998) and Griffies (1998). The discretization of the variational approach is, however, different from the triad approach of Griffies et al. (1998) and uses inherently unstructured grid methods. The Hilbert-space-compatible reconstructions and mimetic differential operators of ICON-O's dynamical core provide a discrete Hilbert space that allows a direct and structure-preserving discretization of the eddy parameterization. Full details are given in Korn (2018). Since the 40km ocean grid used here is barely eddy-permitting, we keep the GM scheme switched on using a default eddy diffusivity parameter of 400 m$^2 s^{-1}$.

As equations of state that approximates the density as a function of potential temperature, salinity and depth we use the UNESCO-80 formulation. For the parameterization of turbulent vertical mixing, ICON-O offers different choices: a Richardson-number-dependent parameterization (Pacanowski & Philander, 1981) (PP) including an additional wind-mixing formulation as in MPIOM, the KPP scheme (Large et al., 1994), or, as the standard setting used here, a scheme based on a prognostic equation for turbulent kinetic energy (TKE) that implements the closure suggested by Gaspar et al. (1990). Vertical dissipation and vertical diffusion are discretized implicitly. The transport of the

oceanic tracers potential temperature and salinity uses a flux-corrected transport method with a Zalessak limiter that utilizes flux calculation by compatible reconstructions (Korn, 2017).

The sea-ice model consists of a dynamic and a thermodynamic component. The thermodynamics of sea-ice describe the freezing and melting of sea-ice by a single-category, zero-layer formulation (Semtner, 1976). The sea-ice dynamics are based on the sea-ice dynamics component of the Finite Element Sea Ice Model (FESIM, (Danilov et al., 2016)), which uses the standard elastic-viscous-plastic (EVP) formulation. As ICON-O applies an analogue of an Arakawa C-grid and FESIM uses an Arakawa A-grid type staggering, an interpolation between the ICON-O grid and the FESIM sea-ice dynamics is necessary. Furthermore, an additional rotation of the oceanic and atmospheric variables is required, because ICON-O uses local coordinates, whereas FESIM is based on rotated geographic spherical coordinates. Besides the computational overhead, the coupling between FESIM and ICON-O introduces numerical diffusion, e.g. at least three grid cell wide passages are required to allow a sea-ice transport. To overcome this limitations we are currently working on the integration of a newly developed sea ice dynamic model (Mehlmann & Korn, 2021).

ICON-O's time stepping applies a semi-implicit Adams-Bashford-2 scheme. The free surface equation is solved implicitly in time with an iterative solver based on the conjugated gradient method. The remaining state variables are discretized explicitly. For details we refer to Korn (2017).

2.3 Ocean Biogeochemistry

In ICON-ESM, ocean biogeochemistry is represented by the HAMburg Ocean Carbon Cycle module, HAMOCC6, which simulates biogeochemical tracers in the water column and in the upper sediment (Ilyina et al., 2013; Paulsen et al., 2017; Mauritsen et al., 2019). In the water column, currently at least 20 biogeochemical tracers are prognostically calculated, generally following an extended nutrient, phytoplankton, zooplankton, and detritus (NPZD) approach, also including dissolved organic matter, as described in Six and Maier-Reimer (1996). The co-limiting nutrients consist of phosphate, nitrate, silicate and iron. A fixed stoichiometry for all organic compounds is considered. Phytoplankton is represented by bulk phytoplankton and diazotrophs (nitrogen fixers; Paulsen et al. (2017)). Particulate organic matter (POM) is produced by zooplankton grazing on bulk phytoplankton and enters the detritus pool. Export production is separated explicitly into $CaCO_3$ and opal particles, each sinking with its own sinking velocity. The POM sinking speed can be assigned using one of the three implemented methods: constant speed, linearly increasing speed with depths below the euphotic zone (also known as the 'Martin curve'; Martin et al. (1987)) or calculated using the recently developed M4AGO scheme (Maerz et al., 2020). The remineralization of detritus throughout the water column is either aerobic (if seawater oxygen concentration $> 0.5\ \mu mol L^{-1}$) or anaerobic by denitrification and sulphate reduction. The upper sediment is resolved by 12 biologically active layers and a burial layer and simulates the dissolution and decomposition of particulate inorganic and organic matter and the diffusion of pore water constituents. The HAMOCC model is also part of the MPI-ESM and has been extensively evaluated in previous single-model, e.g. Ilyina et al. (2013); Paulsen et al. (2017); Müller et al. (2018a); Mauritsen et al. (2019); Maerz et al. (2020) and multi-model studies, e.g. Bopp et al. (2013); Kwiatkowski et al. (2020); Séférian et al. (2020).

Within the HAMOCC core subroutines, only the biological and chemical sources and sinks, as well as tracer sinking and ascending are computed. Therefore, when implementing HAMOCC6 (a model version used in CMIP6 simulations) in ICON-ESM, the HAMOCC6 interface to the ocean and atmosphere components was adjusted to the ICON-ESM infrastructure accordingly. This adjustment includes the transport of bio-

geochemical tracers with the same routines and numerical schemes as the physical tracers of the ICON-O model. As in previous model versions, it was ensured that all chemical constituents in HAMOCC are mass conserving within computational precision in this implementation.

2.4 Land

ICON-Land is a novel framework developed at MPI-M for the modeling of land processes in ICON that clearly separates model infrastructure from land surface process descriptions. It features a flexible scheme of land surface tiling and object-oriented organization of physical and biogeochemical processes. Apart from the ICON-ESM configuration, ICON-Land is used in the ICON-A atmosphere configuration and can also be run in a land stand-alone mode (see e.g. Nabel et al., 2020). The ICON-Land implementation used in the ICON-ESM v1, comprises physical and biogeochemical processes provided by the JSBACH 4 land model, a port of JSBACH 3.2 (Reick et al., 2021) to the ICON-Land framework. Previous JSBACH versions have been the land components of the MPI-ESM versions used in CMIP5 (Giorgetta et al., 2013) and CMIP6 (Mauritsen et al., 2019).

Comparing to Reick et al. (2021) JSBACH 4 features certain improvements of the physical processes at and below the surface, including a five-layer snow scheme and the phase change of water within the soil (Ekici et al., 2014; de Vrese et al., 2021). Also included are the options to calculate the soil thermophysical properties depending on the soil water content and the general properties depending on the organic matter content of a given soil layer. Surface runoff and sub-surface drainage from ICON grid cells are routed through a hydrologic discharge model (Hagemann & Dümenil, 1997) using a novel method for generating river directions (Riddick, 2021); the resulting river discharge is coupled as freshwater flux to the ocean via the YAC coupler (see section 2.5). Surface temperature of lakes is computed by a simple mixed-layer scheme including ice and snow on lakes (Roeckner et al., 2003). The surface energy balance and the soil thermal layers on land are coupled implicitly to the vertical diffusion scheme of ICON-A.

Biogeochemical processes in JSBACH 4 in this study are simplified relative to JSBACH 3 (Reick et al., 2021). Natural vegetation dynamics (Brovkin et al., 2009) coupled to land-use transitions (Reick et al., 2021), as well as the coupling of terrestrial carbon and nitrogen cycle (Goll et al., 2017) have not yet been ported from JSBACH 3, however, are partially planned to be ported for future ICON-ESM versions. In the piControl and historical simulation ensemble (section 4), natural vegetation and anthropogenic land cover change have been prescribed by annual maps of cover fractions on these 11 PFTs based on Pongratz et al. (2008) and transient crop and pasture fractions derived from LUH2 v2h (Hurtt et al., 2019) as described in Mauritsen et al. (2019).

2.5 Coupling

Ocean and atmosphere processes run concurrently and perform a parallel neighbourhood search and data exchange between the two horizontal grids via the YAC coupling library (Hanke et al., 2016). Fig 1. depicts a schematic view of the model components and the exchange of coupling fields.

Here, we use YAC version 1.5, which contains bug fixes and performance improvements. The components of the wind- and velocity vectors are interpolated using Bernstein-Bézier polynomials following Liu and Schumaker (1996). We use the interpolation stack of YAC and fill all target cells, which do not get any data with this standard interpolation, by applying a 4-nearest-neighbour arithmetic average interpolation. The river discharge is remapped to the target grid in a way that each source cell containing a river discharge value is assigned to a coastal target cell on the ocean grid. All other fields are

interpolated using 1st-order conservative remapping. The grids and masks are constructed in a way that all source cells are covered with this standard interpolation. The calculation of the neighbourhood-relations and the interpolation stencils is performed by YAC repeatedly during the initialisation of each model run, solely based on geographical locations of grid cell vertices and centers.

The atmosphere component provides the zonal and meridional components of the wind-stress separately over ice and over water, the surface fresh water flux as rain and snow over the whole grid cell and evaporation over the ocean fraction of the cell, short- and longwave radiation and latent and sensible heat fluxes over the ocean, sea ice surface and bottom melt potentials, the 10 m wind speed and sea level pressure. The ocean provides the sea surface temperature, the zonal and meridional components of velocity at the sea surface as well as ice- and snow thickness, and ice concentration. The data exchange encompasses aggregation, averaging and re-partitioning of the exchange fields. YAC routines are called at every model time step, and data are accumulated inside the YAC library. At user-defined coupling events - here every 1800 seconds - data are averaged and sent to the respective receiving processes.

2.6 Computational configuration and performance

All simulations were performed with the bullx DLC 720 high performance computing system for Earth system research (HLRE-3) of the "Deutsches Klimarechenzentrum" (DKRZ). The simulations utilize 120 "Broadwell" compute nodes of the system named "Mistral", which include 36 processing units each.

The domain decomposition is performed separately for ocean and atmosphere: the ocean decomposition at 40 km horizontal resolution, which includes only ocean grid points, and the global atmosphere decomposition of 158 km horizontal resolution. The sea-ice model is included in the ocean code and runs on the ocean grid, the land model works within the atmospheric decomposition. The YAC coupling library is linked to each of the two components of the ICON model (atmosphere/land and ocean/sea-ice) and performs the aggregation, averaging and re-partitioning of the exchange fields using their respective decomposition.

Due to this technical setup, the load balancing has to be optimized for these two major components, only. The heavy workload due to multiple tracers in configurations with HAMOCC requires different weightings for run with and without ocean biogeochemistry. The best compromise between shortest return time and parallelization overhead was obtained for a load balancing of 74 nodes (2664 mpi-processes) for running ocean and sea-ice (without HAMOCC) on the 40 km grid and concurrently using 46 nodes (1656 mpi-processes) for atmosphere and land on the 158 km grid. With this configuration we achieve an average performance of ten simulation years in one batch-job executing within roughly two hours. Without any queuing-time at the machine (depending on the load of the machine, or by assigning high-priority to the job-chain) it results in a performance of 120 simulated years per day. A hybrid configuration using mpi- as well as openmp- (shared memory) parallelization was tested and exhibited less performance on the DKRZ machine, which is probably due to partly missing optimizations in the code. In the runs including HAMOCC, the best optimization was achieved for a load balancing of 46 nodes for atmosphere and land and 154 nodes for the ocean (physics and biogeochemistry) and sea ice. An average performance of 40 simulated years per day was achieved with this configuration.

3 Tuning and spin-up

3.1 Tuning principles and targets

Model tuning is an integral part of the model development process (Mauritsen et al., 2012). In the coupled system, a first-order tuning goal is to achieve stable climate conditions to minimize drifts in the piControl climate used as reference for climate change simulations. A near-zero top-of-atmosphere energy-flux balance is required as well as long-term stable circulation, for example the Atlantic Meridional Overturning Circulation (AMOC) in the ocean. Furthermore, it is desired to match the model results with observed climate conditions, for example a global mean surface air temperature close to the observational estimates for the second half of the 19th century. Based on experience gained in the tuning of the stand-alone ocean and atmosphere set-ups (Giorgetta et al., 2018), a small number of parameters associated with the parameterization of specific processes were selected for tuning. In the atmosphere these are mainly related to cloud properties and to subgrid-scale orographic processes. Parameters modified for tuning purposes in the ocean include the value of biharmonic viscosity, parameters in the TKE mixing scheme, and the isoneutral diffusion and eddy diffusivity coefficients in the Gent-McWilliams closure (Korn, 2018). Apart from albedo settings in the atmosphere, the sea-ice tuning parameters determine the change in ice-thickness distribution during melting and freezing, respectively (Notz et al., 2013; Mauritsen et al., 2012).

The tuning parameters used in the initial sensitivity experiment and the final spin-up are listed in Table 1.

Table 1. Parameters used for tuning.

ICON Parameter	Symbol in reference	Reference	Description
Atmosphere			
crs	$r_{0,surf}$	Giorgetta et al. (2018)	critical relative humidity for condensation (surface)
crt	$r_{0,top}$	Giorgetta et al. (2018)	critical relative humidity for condensation (upper troposphere)
entrpen	ϵ^1	Nordeng (1994)	entrainment in deep convection
gkdrag	G	Lott (1999)	subgrid-scale orographic (SSO) gravity wave drag
gkwake	C_d	Lott (1999)	SSO low-level blocking
gklift	C_1	Lott (1999)	SSO lifting
Ocean			
K	K	Korn (2018)	GM isoneutral diffusion
κ	κ	Korn (2018)	eddy diffusivity
sea ice			
leadclose 1	cmelt	Notz et al. (2013)	ice area change during melting
leadclose 2/3	cfreeze	Notz et al. (2013)	ice area change during freezing

3.2 Spin-up and tuning history

The ocean initial conditions for temperature and salinity were taken from the Polar Science Center Hydrographic Climatology data set (Steele et al., 2001). First, a 200-year long stand-alone ICON-O simulation was carried out using the atmospheric climatology forcing and the respective bulk formulae described in Marsland et al. (2003). Starting from the restart fields obtained from the stand-alone ocean simulation, several experiments with different tuning choices were conducted, partly sequentially with parameter changes on the fly, partly in parallel to study difference in drift behavior. The final tuning sequence is documented in Fig. 2.

The start of the coupled simulation (slo1304) is characterized by a large drift, where both atmosphere and ocean are cooling and the TOA radiation balance is negative (not shown). Introducing background tropospheric aerosols in the run slo1307 led to even stronger

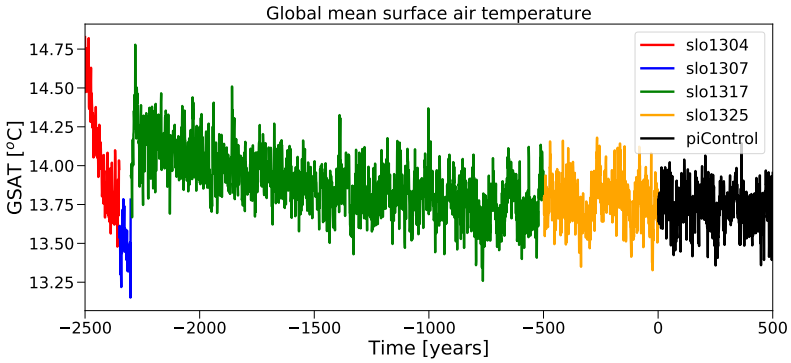

Figure 2. Spin-up history of the coupled simulation: time series of global mean surface air temperature from a sequence of simulations leading to the piControl experiment.

decrease in GSAT that required counter-tuning. This was achieved by increasing the critical relative humidities for condensation *crs* and *crt* and the entrainment parameter *entrpen* which led in particular to a reduction of global mean cloud fraction and an increase of net incoming radiation at the TOA. While this resulted initially in overly strong warming, the long-term drift cooled the model to acceptable values and we finally obtained a solution with small overall drift. While this may appear as a very straightforward tuning process, a large number (order 100) of further model experiments with different parameter settings and simulated lengths of a few years to several hundreds of years have been performed to arrive at this spin-up sequence. The influences of some of the tuning choices on the simulated climate are discussed in several parts of the model evaluation of Section 4 and in cection 5.

A bug-fix related to erroneous snow accumulation in a few grid points required an update of the code in run slo1325 without noticeable effects on the climate. This simulation was carried out for another 500 years, where we defined the start of the piControl simulation. The starting point of piControl was also used to initialize the DECK experiments 1pctCO2 and abrupt4xCO2, and one realization of the historical simulations. Further realizations were started from different dates of piControl. In addition, an AMIP simulation was included using the same code version as the coupled experiments.

After the completion of the DECK experiments presented in this paper, a coding error was detected in the vertical diffusion of ICON-A. The bug is related to the way ocean currents are taken into account in the wind-stress calculation. The effects of the error turned out to be time-step and grid-size dependent and had detrimental effects in a very high-resolution (5km) coupled proto-type model. At the low resolutions applied here, we were able to identify typical effects of this error (e.g. some changes in the representation of the equatorial current system in the ocean), but most of the analyzes and illustrations presented here remain largely unaffected. The most notable effect is an even stronger variance of the ENSO time series, but we diagnosed otherwise very similar characteristics of the variability (not shown). We have concluded that the bug must be fixed but changes to the results were too minor to justify a repetition of the DECK experiments and their post-processing. All conclusions regarding benefits and shortcomings of the ICON-ESM DECK simulations presented here remain unaffected.

The tuning of the ocean biogeochemistry was carried out after the tuning of the coupled setup. A first-order tuning goal for the ocean biogeochemistry in an ESM is to limit significant drifts in the biogeochemical tracer fields and fluxes in the piControl run. Furthermore, parameters are adapted within a reasonable range to drive the model closer to observations. In the first step of the tuning procedure, presented here, the focus is on

the upper ocean biogeochemical fields reflecting the representation of the chemical and biological processes in the model. The initial conditions for the biogeochemical tracer fields in the water column and sediment were interpolated from a previously well spun-up MPIOM piControl run. Ocean and atmosphere were initialized from the end of the slo1325 run and ICON-ESM was run with the piControl climate. The atmospheric CO_2 concentration was set to 278 ppm, representing a pre-industrial climate. The dust deposition climatology of Mahowald et al. (2005) and historical nitrogen deposition fields from the CMIP6 input database (https://esgf-node.llnl.gov/projects/input4mips/) were used. The POM sinking speed was calculated based on the Martin curve.

To account for the ocean circulation simulated by ICON-ESM, some of the HAMOCC tuning parameters were changed from their default values. The appropriate weathering rates, which are used to compensate the loss of carbon and nutrients from the water column to the sediment, were calculated and updated during the spin-up procedure. After a simulation length of 500 years, the model approached a semi-steady state in the upper ocean global monitoring values such as the global surface alkalinity, POM export and nutrients.

Table 2. Overview on the ICON-ESM simulations.

Experiment	Description	Period	Ens.size	Initialization
piControl	Preindustrial Control	500 years	1	spin-up run
1pctCO2	idealized CO_2-increase	150 years	1	spin-up run
abrupt4xCO2	idealized CO_2 increase	150 years	1	spin-up run
historical	Transient forcing	1850-2014	5	piControl (yrs 0, 100, 200, 300, 400)
AMIP	atmosphere-only	1978-2014	1	n.a.

4 Model evaluation

The set of experiments described in this paper is listed in table 2. We start with a brief account of the piControl experiment. Since the evaluation in comparison with observations is based on data from the recent decades, we base the analyzes on the historical ensemble. The idealized climate change experiments 1pctCO2 and abrupt4xCO2 are used to estimate the climate sensitivity of ICON-ESM in section 4.4.

4.1 The pre-industrial control simulation (piControl)

The global mean surface air temperature (GSAT) is stable over the 500 year long piControl simulation with a small cooling of -0.01 K per century (Fig. 3). The time mean of 13.73^oC is consistent with estimates of warming over the historical period and the estimated global temperatures from reanalyses (Hawkins & Sutton, 2016). The goal of a very stable TOA radiation is achieved and only a small residual of less than 0.02 Wm^{-2} assures that long-term integrations can be run with this model version. The AMOC is a key quantity for the meridional heat exchange in the Atlantic Ocean and its stability is important for maintaining a proper sea-ice distribution and North Atlantic deep water formation (for more details see section 4.2.5). The control run has a time-mean AMOC strength at 26^oN of slightly less than 16 Sv ($1Sv = 1Sverdrup = 10^6 m^3 s^{-1}$). The AMOC is stable over the simulation, but exhibits relatively strong multidecadal variations with an amplitude of up to 3 Sv.

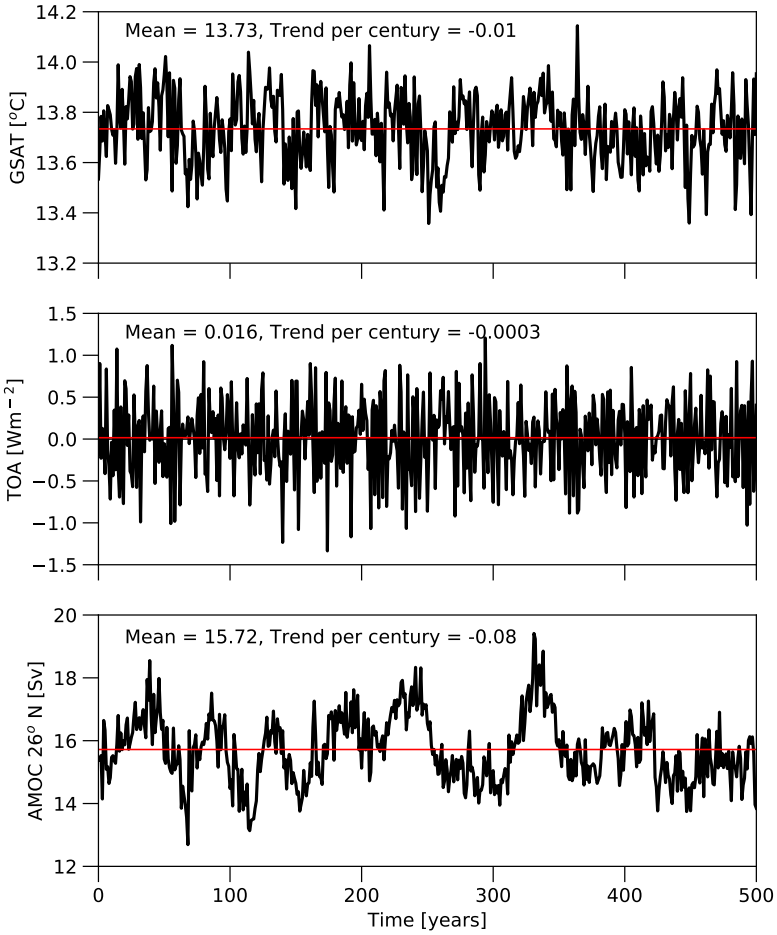

Figure 3. Evolution of key quantities during the piControl experiment: (upper) global mean surface air temperature, (middle) top-of-atmosphere radiation balance, and (lower) the strength of the Atlantic Meridional Overturning streamfunction at 26^oN and 956m depth. Red lines indicate the time mean.

4.2 The historical simulation ensemble

4.2.1 Temperature evolution during the historical period

ICON-ESM reproduces the evolution of the global mean surface temperature largely in good agreement with observational products (Fig. 4). The mid-20^{th} century warming and the subsequent cooling towards the 1970s agree in magnitude and timing, and the effects of volcanic eruptions like Agung (1963/64) and Pinatubo (1991) are captured. The simulations slightly disagree with the observational records in the late 20^{th} to early 21^{st} century because the model overestimates the warming trends from the 1970s onward. On the other hand, the warming trends are underestimated near the end of the simulation so that the simulated temperatures agree with the observations at the end of the simulated period. A decomposition into northern (Fig. 4b) and southern (Fig. 4c) hemispheres reveals that the deviations stem mainly from the northern hemisphere. Assessing the reason for this discrepancy requires further investigations but the more pronounced biases in the northern hemisphere point to an underestimation of the cooling effect of anthropogenic aerosols (Mauritsen et al., 2019) rather than too high climate sensitivity in ICON-ESM (see section 4.4).

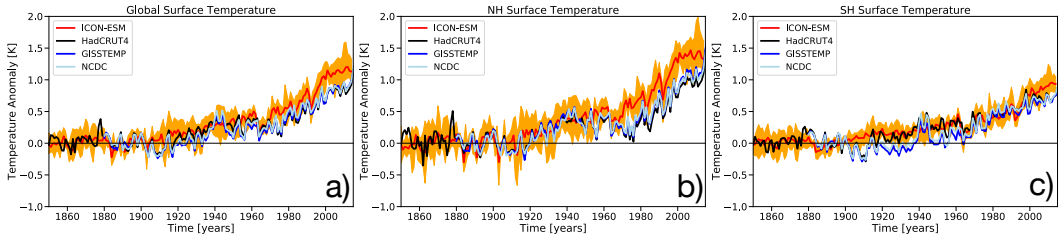

Figure 4. Time series of surface temperature over a) the globe, b) the northern hemisphere, and c) the southern hemisphere for (red-orange) the ICON-ESM historical ensemble, and observational compilations by (blue) the Goddard Institute for Space Studies Surface Temperature product (Lenssen et al., 2019), (black) the blended Hadley Center/Climate Research Unit global temperature data set (Morice et al., 2012), and (light blue) the NOAA NCDC historical merged land–ocean surface temperature data set (Smith et al., 2008; Zhang et al., 2019). The simulated global temperature is constructed using SSTs over the ocean and surface air temperatures over land.

4.2.2 Atmosphere

Table 3. Data used for evaluation of atmospheric quantities. Further data used for the computation of skill scores are specified by Crueger et al. (2018).

Quantity	Name	Period	Reference
sea level pressure	ERA-Interim	1979 - 2014	Dee et al. (2011)
zonal mean temperature	ERA-Interim	1979 - 2014	Dee et al. (2011)
zonal mean zonal wind	ERA-Interim	1979 - 2014	Dee et al. (2011)
cloud fraction	CALIPSO-GOCCP (v3.1.2)	2007-2019	Chepfer et al. (2010)
precipitation	GPCP (v2.2)	1979-2013	Adler et al. (2003)

For the evaluation of atmospheric quantities we follow as closely as possible plotting styles and use of data sets as in Crueger et al. (2018) to enable a comparison of the performance of the coupled ICON-ESM with AMIP-style (i.e. atmosphere-only) simulations by ICON-A and predecessors. Data sets used in the comparison are listed in Table 3. We only use data from after the beginning of the satellite era and compare to ICON-ESM output from the same period of the historical simulation. As model biases are in general large in comparison to the spread between different ensemble members we only use one ensemble member for the comparisons.

To allow a quantitative comparison of global model performance with predecessors and uncoupled simulations of this and earlier model versions, we present skill scores for simulated annual mean quantities as proposed by Reichler and Kim (2008) in Fig. 5. We calculate these scores in the same way and with respect to the same observational data as described by Crueger et al. (2018). Skill scores for model biases are calculated with respect to model biases in a reference simulation for which we use the historical CMIP6 simulation with the MPI-ESM-LR-1.2 (Mauritsen et al., 2019). It is obvious that the ICON-ESM performs worse than the reference model for many quantities both globally and in the three geographic regions tropics, northern and southern extratropics. It performs also

worse than the predecessor MPI-ESM-LR (Stevens et al., 2013). The performance has clearly improved only for some quantities in the southern extratropics.

In general, a better agreement of uncoupled simulations with observations is expected as they are driven by observed sea surface temperatures and sea ice. This better agreement is clearly visible in Fig. 5, where uncoupled scores for many quantities are below one in most regions, i.e. the agreement with observations is better than in the coupled reference simulation. Differences between our simulation and the uncoupled AMIP experiment of Crueger et al. (2018) are expected due to the coupling, but also due to parameter changes related to the tuning of the coupled model as described in Sections 3.2 and small code modifications as described in Section 2.1. The latter two effects can be estimated from comparing the skill scores of the ICON-ESM AMIP simulation with the predecessor ICON-A-1.3.00 used by Crueger et al. (2018). Although some quantities improve, the skill in the uncoupled experiment is lower than in the AMIP simulation. A lower skill can be expected because our tuning aimed at the performance in the coupled simulation. In the following we evaluate the spatial patterns of some atmospheric quantities.

Figure 6 shows annual mean sea level pressure from the ERA-Interim reanalysis and the difference of the ICON-ESM to this dataset. Maximum anomalies of up to about 10 hPa are of the same order as anomalies simulated in the uncoupled model (Crueger et al., 2018). However, the spatial structure is very different. While in the uncoupled model there was an underestimation in most parts of the tropics and sub-tropics and a strong positive bias in particular over the Arctic, here we simulate strong positive biases centred near about 45° in both hemispheres. Extratropical biases showed some sensitivity to the SSO parameters (see Table 1) in the tuning process. The Arctic positive bias of mean sea-level pressure over the Arctic found in Crueger et al. (2018) could be reduced by activating SSO mountain lift forces of using the parameter *gklift*. The tropical low bias was a feature in all our tuning attempts.

According to the skill scores presented in Fig. 5, the ICON-ESM simulates precipitation over land and ocean in the extratropics similar or even better than predecessors or uncoupled model versions. It performs worse, however, in the tropics. Fig. 7 shows annual mean precipitation patterns in the ICON-ESM in comparison to data from the Global Precipitation Climatology Project (GPCP). The model simulates the typical distribution of tropical and extratropical rainfall patterns. In the Pacific, rainfall maxima are too high and a double-ITCZ bias, typical for many climate models, can be identified. These features were also reported for the uncoupled ICON AMIP simulation from Crueger et al. (2018). Improvements with respect to this uncoupled simulation can be identified in the tropical Atlantic and Indian oceans. Concerning the seasonal cycle, a major bias is a shift of maximum precipitation from summer to winter in the boreal forest zone (50N-65N) over the continental interior of Eurasia (not shown), which leads to a large regional deficit in simulated vegetation productivity.

Figs. 8 and 9 show global annual mean total cloud fraction and zonal mean vertically distributed cloud fraction, respectively, in comparison to the GCM Oriented Cloud Calipso Product (CALIPSO-GOCCP) data. Total cloud fraction is clearly too low in subtropical regions in both hemispheres, a feature which was visible but less strong in the AMIP simulations of Crueger et al. (2018). The vertical distribution of cloud fraction indicates that this is in particular related to an underestimation of low clouds in the subtropics and tropics. High clouds are, by contrast, overestimated in the tropics and middle to high latitudes. Different tuning choices would be able to alleviate these deficiencies, but we did not reach a global energy balance for a realistic global mean temperature and better cloud distributions at the same time.

The skill scores indicate that the performance of the ICON-ESM for annual zonal means of both temperature and zonal wind is in general worse than that of its prede-

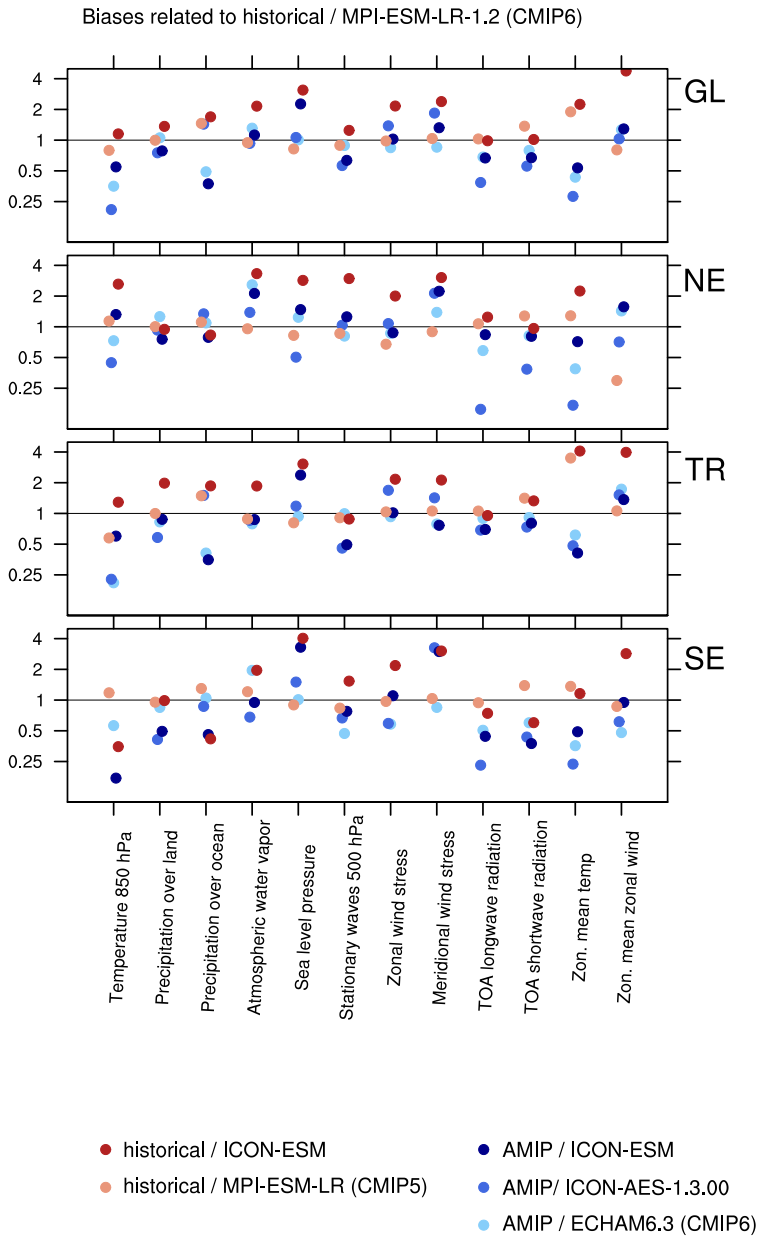

Figure 5. Standardized annual mean climatological errors of selected variables in several simulations with reference to the CMIP6 historical simulation with MPI-ESM-LR-1.2. A value smaller/larger than 1 indicates a smaller/larger bias compared to this reference for the evaluation period 1979 – 2008. Scores are averaged over (from top to bottom) the full globe, the northern extratropics, the tropics (30°S - 30°N), and the southern extratropics. Colored dots indicate scores for the coupled simulations with (red) the ICON-ESM and (orange) the MPI-ESM-LR (Stevens et al., 2013), as well as for the AMIP simulations with (dark blue to light blue) the ICON-ESM, ICON-A-1.3.00 (Crueger et al., 2018), and ECHAM6.3 (Mauritsen et al., 2019). See Section 4.2.2 for further details on the skill scores.

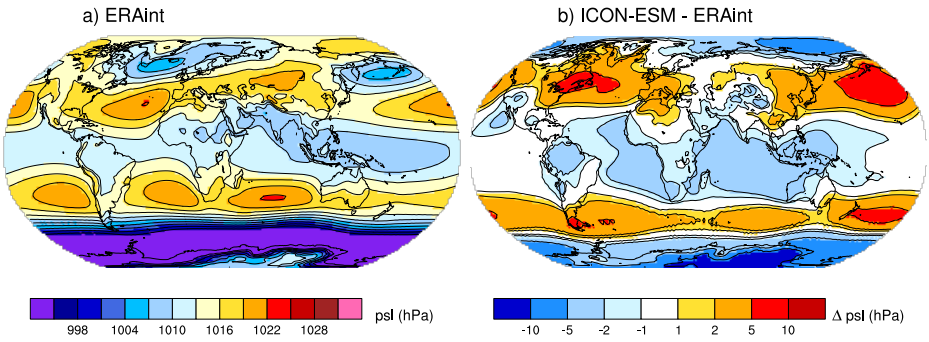

Figure 6. Sea level pressure (hPa) averaged over the period 1979-2014 a) from the ERA-Interim reanalysis and b) difference between ICON-ESM historical simulation and ERA-Interim.

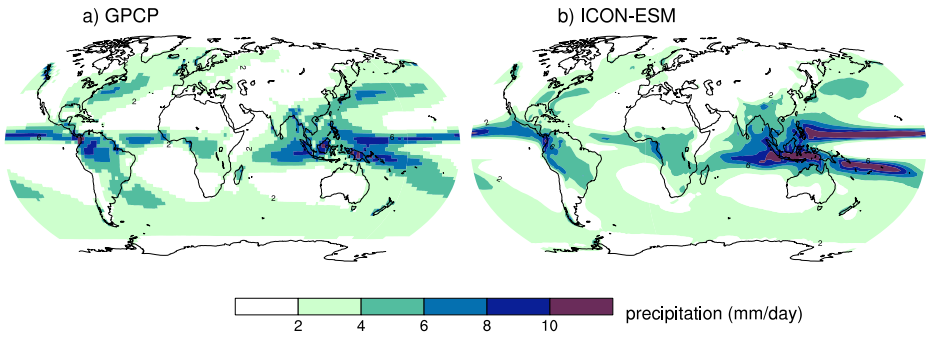

Figure 7. Precipitation (mm/day) a) from the GPCP observations averaged over 1979-2013 and b) from the ICON-ESM historical simulation averaged over 1979-2014 and GPCP.

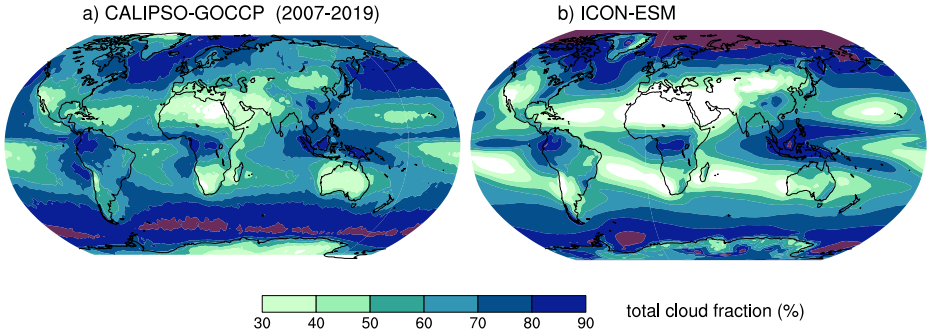

Figure 8. Total cloud fraction (%) a) from the CALIPSO-GOCCP observations averaged over 2007-2019 and b) from the ICON-ESM historical simulation averaged over 1979-2014.

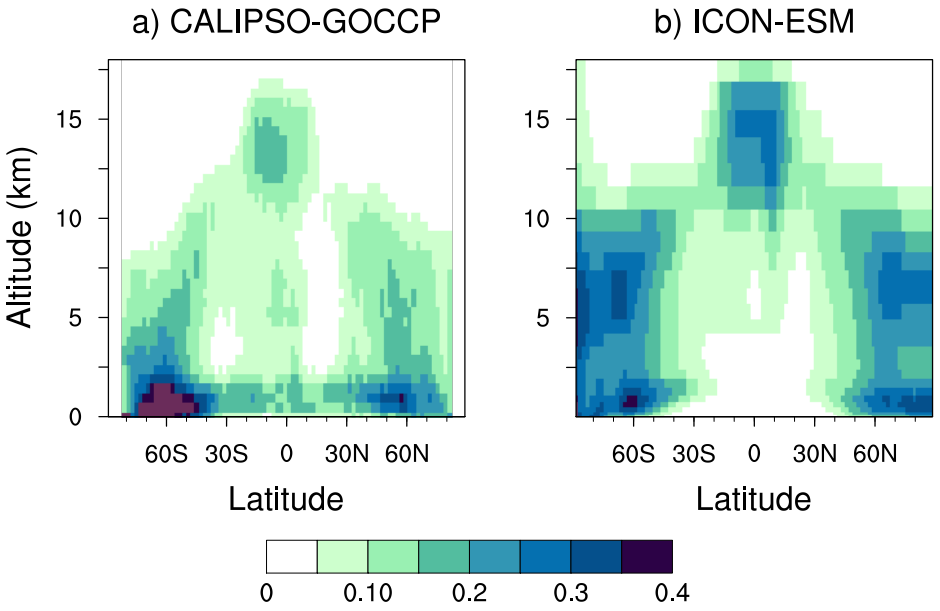

Figure 9. Zonal mean vertical distribution of cloud fraction (%) a) from the CALIPSO-GOCCP observations averaged over 2007-2019 and b) from the ICON-ESM historical simulation averaged over 1979-2014.

cessors. In particular, the score for zonal wind is high in all geographical regions, but it should be noted, that zonal wind biases were very low for the reference simulation (Fig. B3, Mauritsen et al., 2019). The large positive temperature bias in the high latitude middle atmosphere and the cold bias near the high-latitude tropopause (Fig. 10) are recurrent features of our models (Crueger et al., 2018), but partly stronger in the ICON-ESM. In the troposphere, the model shows in general a warm bias at low and a cold bias at high latitudes. The large zonal wind bias (Fig. 11) is dominated by too strong westerlies in the mid-latitude troposphere and stratosphere, a feature which is strongest in both hemispheres during boreal winter. The position of subtropical jets is biased poleward in both hemispheres. A reduction of the zonal wind biases would be possible through different tuning choices in the parameterization of SSO effects, but in our tuning experiments this came in general at the expense of larger biases in sea ice and the AMOC.

4.2.3 Land

We compare our model ensemble results for the surface albedo with the MODIS MCD43C3 CMG Albedo Product (C. Schaaf & Wang, 2015). Cescatti et al. (2012) and C. B. Schaaf et al. (2002) show that the product is suitable for climate model comparisons. It comes with quality information for each data point (quality flags). These flags condense uncertainties in the elicitation of the data, such as atmospheric scattering and absorption, anisotropy, inadequate temporal, spatial and spectral sampling, and narrowband to broadband conversions. For our comparison we first exclude MODIS data with a minor quality of the inversion (quality flags 4 and 5). Then we interpolate the data from the original MODIS grid of 0.05^ox 0.05^o (about 5.6 km at the equator) and from our model grid to a Gaussian lon-lat grid of 96 x 192 (about 1.88^o or 210 km at the equator). As the albedo varies strongly throughout the year due to variations in the angle of the incoming radiation, in leaf area index (LAI), and in snow cover, we take January and July data to represent the winter and summer extremes, for which we average our

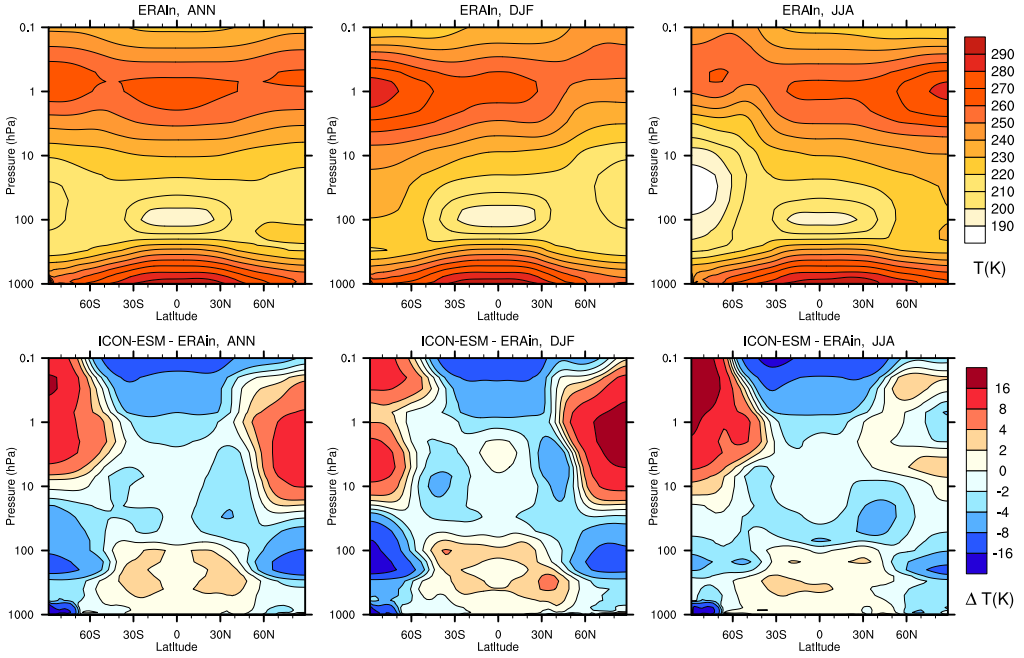

Figure 10. Zonal mean temperature (K) averaged over the period 1979-2014 (top row) from the ERA-Interim reanalysis and (bottom row) difference between ICON-ESM historical simulation and ERA-Interim. From left to right are shown the annual, boreal winter (DJF), and austral winter (JJA) averages.

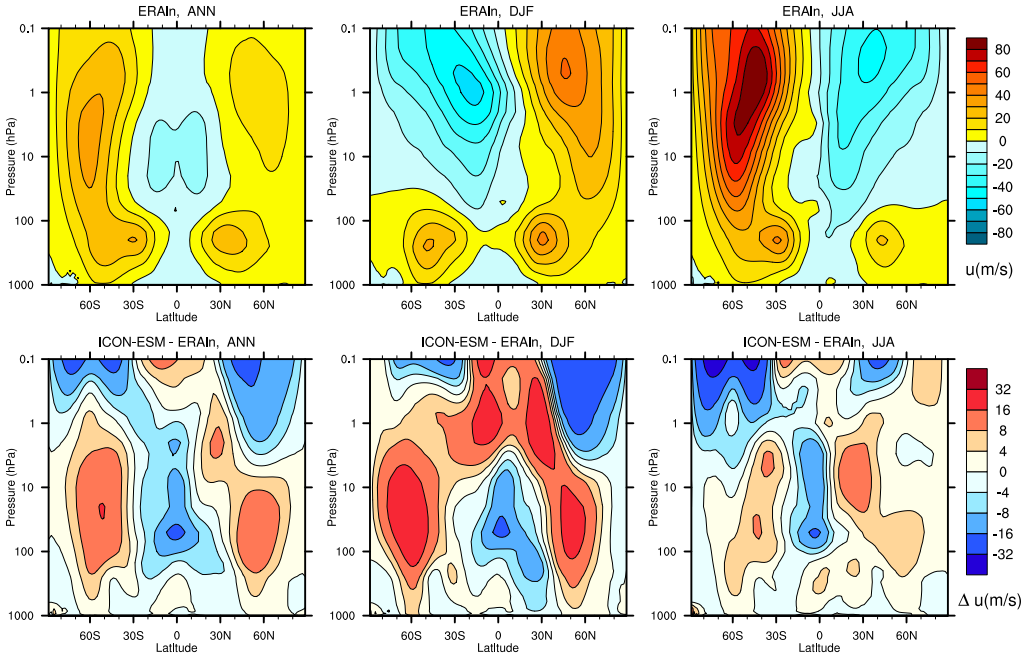

Figure 11. Zonal mean zonal wind (K) averaged over the period 1979-2014 (top row) from the ERA-Interim reanalysis and (bottom row) difference between ICON-ESM historical simulation and ERA-Interim. From left to right are shown the annual, boreal winter (DJF), and austral winter (JJA) averages.

model results and the MODIS data over the years 2001 till 2014. The differences are shown in Figure 12.

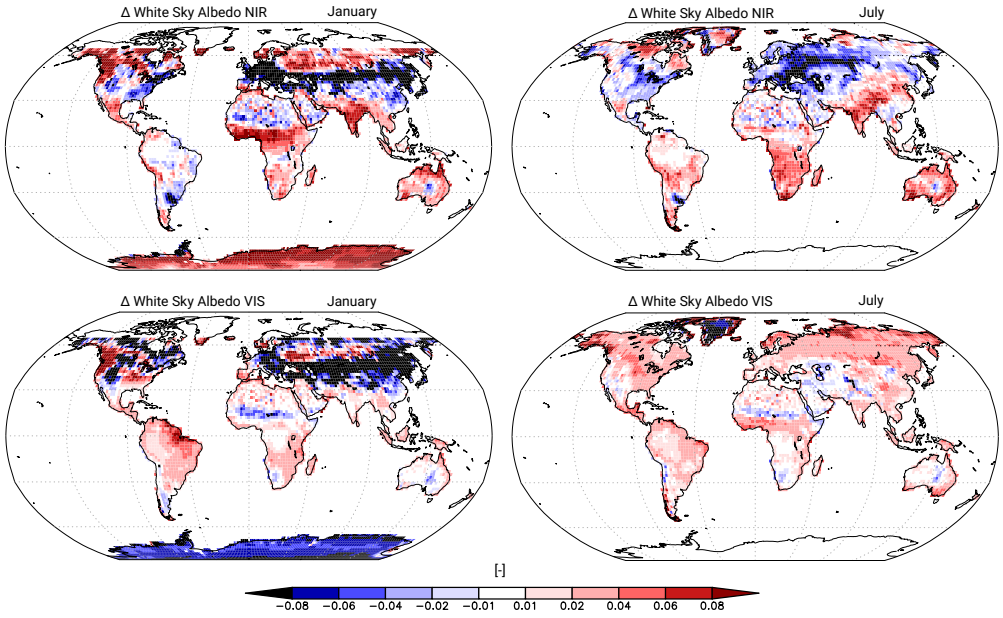

Figure 12. White Sky Albedo (WSA) differences between the ICON-ESM historical ensemble and the MODIS data. Shown are the NIR and VIS bands for January and July averaged from 2001 till 2014.

All albedo differences are in the range +/- 0.1. In general the biases are weak as compared to the absolute MODIS albedos. E.g. in January the global NIR albedo is 0.31 for the absolute values of MODIS while the corresponding bias is only 0.003. Over glaciers we find a common pattern where the near-infrared (NIR) albedo is too high and the visible (VIS) albedo is too low (see in January over Antarctica and in July over Greenland), which is a direct result of the prescribed minimum and maximum albedo values for glaciers in JSBACH 4. In January, NIR and VIS albedo are too low in the northern mid latitudes, especially in eastern Europe and central Asia. Further analysis reveals that these biases are largely caused by a too small snow cover in JSBACH 4 (not shown). In July, the NIR albedo in eastern North America and large parts of Asia is too low. These low albedos are caused solely by the prescribed soil albedo of the model. Except for the mentioned areas, the albedos tend to be higher in JSBACH 4, e.g. in most of Africa, Australia and India. However, the causes for this overestimation are rather complex and their investigation is beyond the scope of this paper; a deeper analysis of this issue will be published in a forthcoming paper.

For the evaluation of land surface temperature (LST) of our model ensemble, we use the MOD11C1 Moderate Resolution Imaging Spectroradiometer (MODIS) Terra Land Surface Temperature/Emissivity V006 data set (Z. Wan et al., 2015). For our analysis we excluded the data points where the quality flags indicate no retrieval because of clouds. The spatial resolution of the data set is 0.05^o x 0.05^o (about 5.6 km at the equator). Furthermore, we compare our results with the first five ensemble members of the MPI-ESM CMIP6 historical simulations (Wieners et al., 2019).

The global ICON-ESM LST time series (Figure 13, green lines) fits well with simulation results from MPI-ESM 1.2 (blue lines) but shows a higher monthly variability

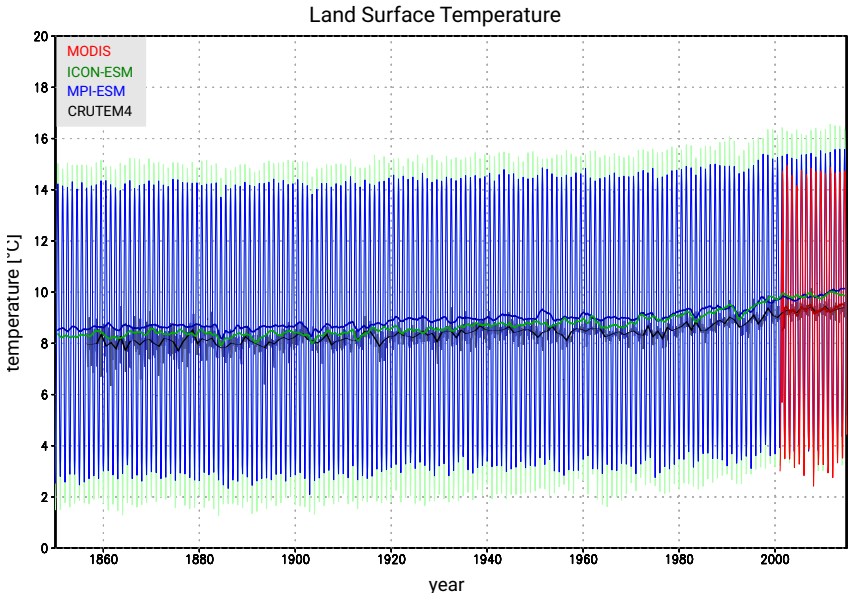

Figure 13. Land Surface Temperature evolution over historical times. Thin lines are monthly means, thick lines are yearly means. Red: MODIS data. Green: ICON-ESM historical ensemble. Blue: MPI-ESM CMIP6 ensemble mean of the first five ensemble members. Black: CRUTEM4 historical reconstruction. Note, as CRUTEM4 includes only temperature changes without a determined absolute temperature the absolute offset is chosen to fit the curve with the satellite data.

(about 2^{o}C). Over the historical period until about 1990, the annual averages of ICON-ESM are slightly colder than those of MPI-ESM 1.2 but about 0.6^{o}C warmer than observed by MODIS (red lines) while monthly variability is - as for MPI-ESM - about 2^{o}C higher than observed. However, its January temperatures are more or less the same as for MODIS, only the July temperatures are much higher causing the warmer annual temperatures. Even when on short time scales both models annual means are not in good agreement with the historical CRUTEM4 reconstruction (black lines), they agree with long term trend (e.g. the temperature rise after 1980).

The geographical LST comparison between ICON-ESM and MODIS averaged between 2001 and 2014 (Figure 14) reveals strong regional differences. The zonal means show a warm bias in the inner tropics and the extratropics. In principle, this can also be seen in the January and July averages. Regionally, the warm bias is throughout the year mostly pronounced in Europe, central Asia, central to north-eastern N-America, the Amazon region, and western Antarctica. The warm bias over Eurasia in January is at least partly caused by the too low snow cover and the associated snow-albedo feedback. The comparison with MPI-ESM (Figure 15) shows that the warm bias in Europe, central Asia and Amazonia, which is most pronounced in July, was much weaker or not existent in MPI-ESM. ICON-ESM is colder than the MODIS data in the outer tropics (see zonal means). Regionally, the locations of the cold biases vary with the seasons but Australia, India and central eastern Antarctica (except for the coastal areas) are colder throughout the year. In Australia and central eastern Antarctica the cold bias was much weaker or not existent in MPI-ESM.

Overall, the zonal mean bias pattern of ICON-ESM as compared to MODIS depicts more or less the global atmospheric circulation and thus indicates mainly an atmospheric origin. Due to the complex continental distribution a land origin seems un-

likely. Nevertheless, the albedo biases surely contribute to the LST biases of ICON-ESM, especially in central Asia and over glaciers.

4.2.4 Ocean

The simulated sea surface temperature (SST) obtained from the ensemble mean of the historical simulations and averaged over the period 1980 – 2014 is compared to the PHC 3.0 data set in Figure 16 a, c. Deviations from the observation-based data set are largely smaller than 1^oC over the open oceans, but we diagnose prominent regions with large errors. The most pronounced cold anomalies are found in the subpolar North Atlantic. As in many other coarse-resolution and even eddy-permitting models (e.g., Keeley et al., 2012) this feature is related to an overly zonal North Atlantic Current (Drews et al., 2015) and likely also related to too weak meridional heat transport. For MPI-ESM, Gutjahr et al. (2019) have shown that moving to eddy-resolving resolution in the ocean improves the sub-polar cold bias in the North Atlantic. Corresponding errors in atmospheric sea-level pressure (Fig. 6) indicate that wind-driven circulation biases also contribute to the error. Large atmospheric SLP and circulation biases in the North Pacific are likely responsible for strong warm biases over the Kuroshio region.

The cold-tongue bias in the equatorial Pacific is also a well known model feature, but the ICON-ESM performs less well than the MPI-ESM (Müller et al., 2018b), even though MPI-ESM-LR features much lower resolution in the ocean compared to the ICON ocean used here. The bias is a surface expression of the generally too cold sub-surface waters in the tropical oceans (Fig. 17) so that the outcropping isotherms in the western Pacific are too cold. In the coupled system, the equatorial cold bias is important for the variability characteristics of ENSO and the associated precipitation distribution (section 4.2.7). Warm biases are diagnosed in the upwelling regions at the western coasts of the tropical oceans. They are most pronounced at the African coast south of the Equator. These features are common in coupled models and, in particular in the case of Africa, are related to insufficient resolution in the atmosphere where coastal orography and alongshore winds cannot be properly simulated (Milinski et al., 2016).

The sea surface salinity (SSS) biases (Fig. 16b, d) in ICON-ESM are relatively small over most of the oceans, except the high northern latitudes and around the Antarctic continent. The Arctic fresh bias extends also into the sub-polar North Atlantic, where overly fresh water is transported with the gyre circulation into the interior ocean making the cold bias in Fig. 16c a fresh bias as well.

The time-mean bias in zonal averages over the global ocean reflect misrepresentations of water mass pathways and processes like vertical and along-isopycnal mixing. The most prominent error feature in the ICON-ESM ocean is a generally too cold interior ocean with strong cold biases in the sub-tropical and tropical oceans (Fig. 17) that are accompanied by overly fresh conditions. The reasons for the overly strong cooling are not fully understood, but comparison with an earlier version using the Pacanowski-Philander mixing scheme (PP, Pacanowski and Philander (1981)) showed that the PP scheme in combination with the wind-mixing parameterization used in MPIOM (Marsland et al., 2003) showed a better performance than the TKE scheme used here. Regarding the upper ocean, this is consistent with the findings of Gutjahr et al. (2019) for the MPI-ESM. Since the TKE scheme is more advanced and shall be further improved by including an energy-consistent scheme for the background mixing in the interior (see Gutjahr et al., 2021), we decided to keep the TKE scheme and attend to an improved tuning in forthcoming versions of ICON-ESM.

The overly cold sub-thermocline waters lead to a too strong stratification near the thermocline in particular in the tropical oceans and are likely responsible for too strong ENSO variability (see section 4.2.7). The salinity biases resemble those in temperature in the tropical and southern sub-tropical region, but higher northern latitudes feature

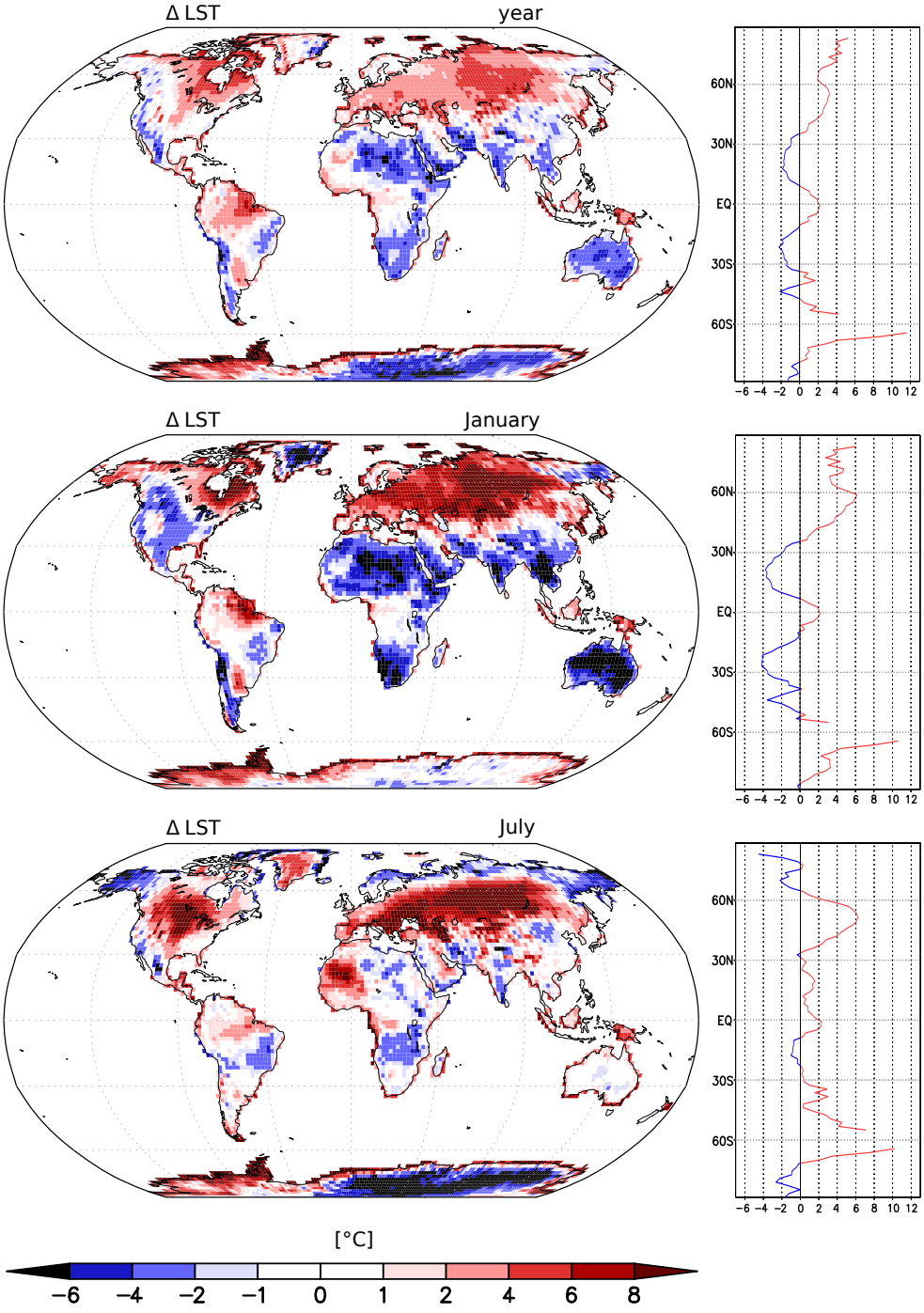

Figure 14. LST differences between the ICON-ESM historical ensemble and MODIS data averaged from 2001 to 2014. The curves on the right side show zonal means.

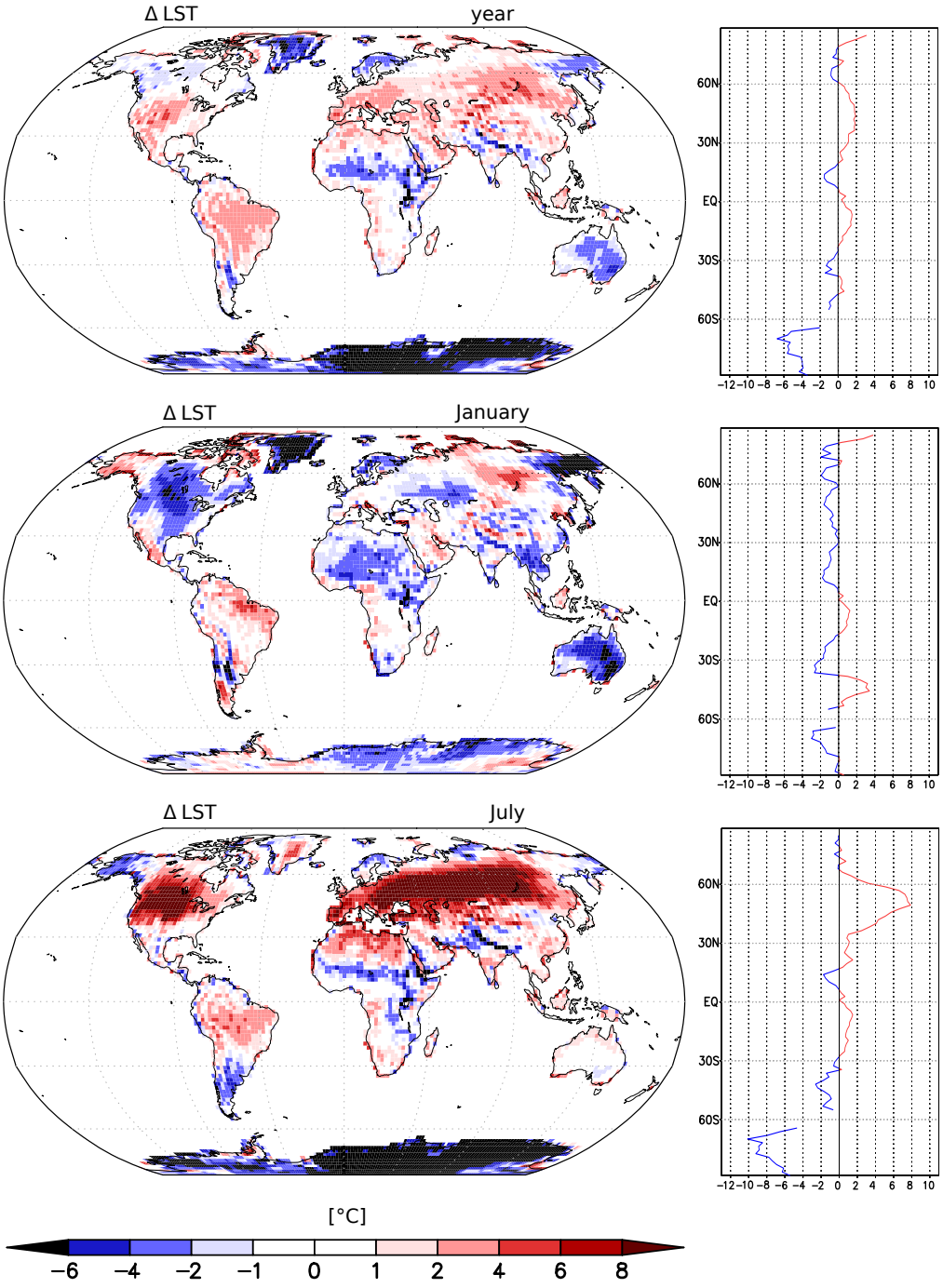

Figure 15. LST differences between the ICON-ESM historical ensemble and MPI-ESM ensemble mean (r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1 and r5i1p1f1) averaged from 2001 to 2014. The curves on the right side show zonal means.

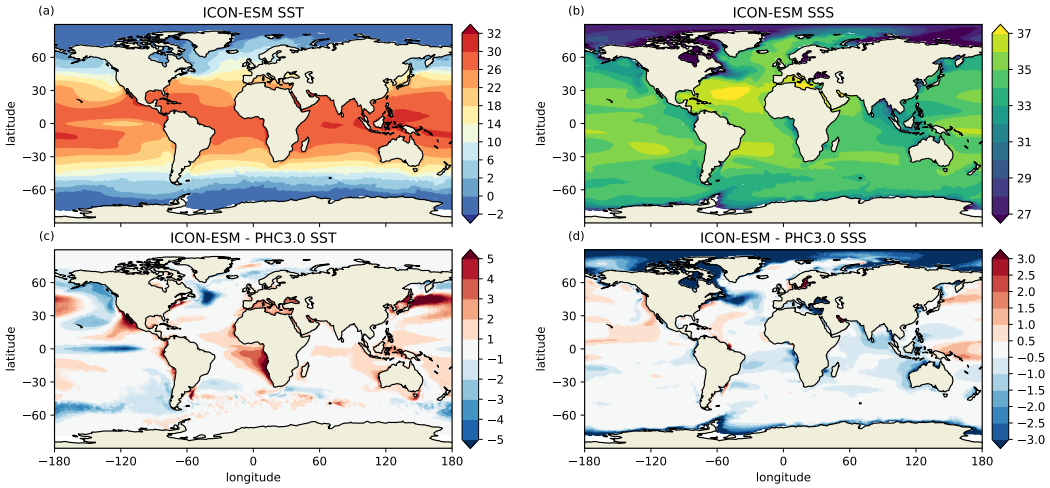

Figure 16. Climatology (1980 -2014) of a) sea surface temperature and b) sea surface salinity simulated in the ICON-ESM historical simulations together with the model biases for c) SST and d) SSS with respect to the observation-based PHC climatology (Steele et al., 2001). TODO: add labels

pronounced positive salinity in the sub-surface ocean. These underlay a too fresh surface layer and lead to a too strong halocline in the Arctic Ocean. The too fresh surface salinities could be related to an underestimation of Fram Strait sea ice export.

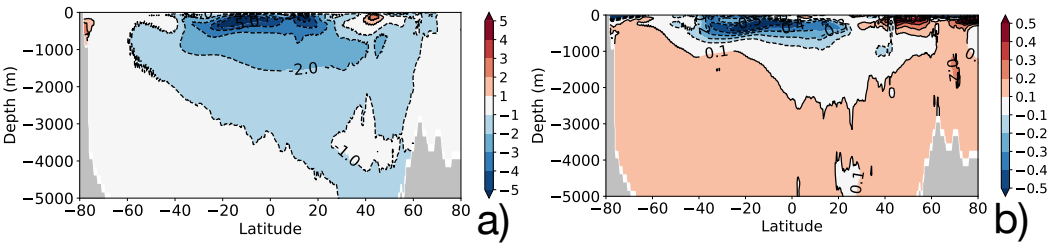

Figure 17. Zonal mean global ocean a) temperature, and b) salinity bias relative to the PHC climatology.

4.2.5 Large-scale ocean circulation

The AMOC is an important part of the global overturning circulation and it is an important carrier of heat and fresh water in the Atlantic. The AMOC stream function (Fig. 18a) represents the zonally integrated view. Facing west, the North Atlantic Deep Water cell is oriented clockwise and includes the downward motion associated with deep water formation in the Labrador Sea and Nordic Seas, as well as the overflows across the Greenland-Scotland Ridge. The maximum strength of the AMOC exceeds 16 Sv at approximately 40^{o}N and we diagnose an export of about 14 Sv at 30^{o}S. The lower, counterclockwise oriented cell is associated with Antarctic Bottom Water (AABW) entering the Atlantic and upwelling in the basins. The strength of the AABW cell in the North Atlantic is roughly consistent with the observations from the RAPID project, but for the South Atlantic the simulations likely underestimate the AABW inflow. While the sim-

ulated AABW maximum is slightly above 2 Sv, observation-based estimates point to values of 6 Sv (Ganachaud & Wunsch, 2003) or 8 Sv (Talley et al., 2003). Although the overturning stream function cannot be compared directly with observation, the RAPID project has provided measurements of the respective flow components at 26.5°N. The profile obtained over more than a decade of observations (Smeed et al., 2018) is shown together with the profiles from the piControl runs and one historical simulation (Fig. 18b) and with their variability range estimated from the standard deviation of monthly fields. In the upper ocean, the model reproduces well the shape of the profile, the maximum near 1000 m depth and also the range of variability. The zero crossing is located above 4000 m in the model, but resides a few hundred meters deeper in the RAPID data. Compared to various versions of MPI-ESM1.2 (Gutjahr et al., 2019) and other CMIP6 models (e.g. Held et al. (2019); Danabasoglu et al. (2020)), which show even shallower NADW cells, the ICON-ESM results are, however, closer to the observations.

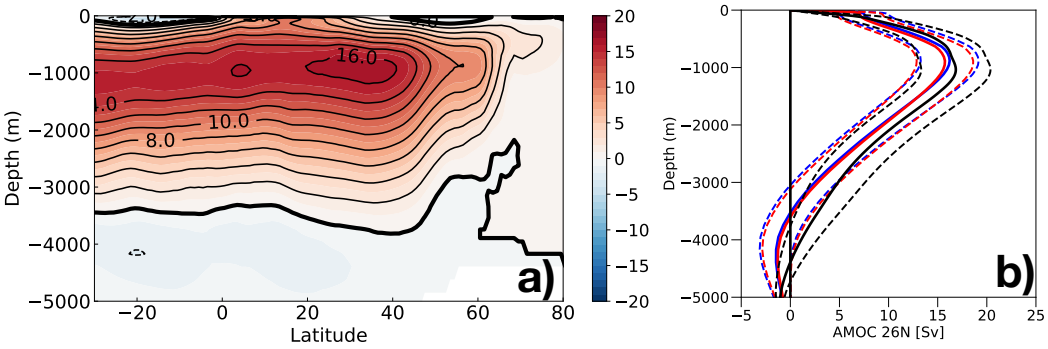

Figure 18. AMOC in ICON-ESM: a) ensemble mean overturning stream function in the Atlantic from the five historical simulations averaged over the period 1980-2014; b) AMOC profile at 26.5N for the historical ensemble (blue) and the piControl simulation (red) together with the observational estimate from RAPID (black); shown are the mean profiles (thick lines) and the range of variability derived from monthly standard deviations.

Table 4. Simulated and observed net volume transports (Sverdrups) across sections (positive means northward).

Section	ICON-ESM	Obs.	Reference
Bering Strait	0.7-0.8	0.7-1.1	Woodgate et al. (2006, 2012)
Fram Strait	- (2.6-2.8)	-1.75 $\pm$5.01	Fieg et al. (2010)
Denmark Strait	- (4.7-4.9)	-4.8	Hansen et al. (2008)
		-3.4 $\pm$ 1.4	Jochumsen et al. (2012)
Iceland-Scotland	4.7-4.9	4.8	Hansen et al. (2008)
		4.6 $\pm$ 0.25	Rossby and Flagg (2012)
Indonesian Throughflow	9.9-10.1	11.6 - 15.7	Gordon et al. (2010)
Drake Passage	108-112	134.0 $\pm$ 14.0	Nowlin Jr. and Klinck (1986)
		173.3 $\pm$ 10.7	Donohue et al. (2016)

The transport through selected passages reflects the representation of the large-scale ocean circulation and can be evaluated in comparison with observational estimates. The total depth-integrated transports through the passages given in Table 4 are mostly simulated within the observational uncertainty found in the literature. The transport through Bering Strait is a key element of the Arctic fresh water budget and the historical simulations are close to the estimates by Woodgate et al. (2006, 2012). The exchange of water masses between the Atlantic Ocean and the Nordic Seas is important for the overturning circulation. The simulated transports are consistent with the circulation scheme by Hansen et al. (2008). The Indonesian Throughflow is another important contributor to the warm-water route of the global conveyor. The simulated transports are slightly underestimated in comparison with the estimate by Gordon et al. (2010). The Drake Passage transport is clearly underestimated in ICON-ESM, both in comparison with the traditional estimate around 135 Sv (Cunningham et al., 2003; Nowlin Jr. & Klinck, 1986) and with the more recent compilation by Donohue et al. (2016). At this stage it is not clear if the mismatch between observed and simulated Drake Passage (and likely Antarctic Circumpolar Current) is related to biases in the wind stresses or the settings of the eddy and background diffusivity in the ocean circulation (P. R. Gent et al., 2001).

The equatorial hydrography and the representation of the zonally-oriented current systems is important for variability features such as Tropical Instability Waves and for coupled ocean-atmosphere phenomena like the El-Nino Southern Oscillation (ENSO). Johnson et al. (2002) provided a compilation of temperature and salinity data and current measurements from the 1990s, which we compare with ICON-ESM results in Figure 19. At 110^{o}W, the placement of the eastward equatorial undercurrent (EUC) and the westward flows on its northern and southern flanks are reproduced well in the model and the depth of the core of the undercurrent lies roughly at the correct depth of 75m (Figure 19a, b). The model underestimates, however, the speeds of the eastward undercurrent and the westward currents flanking the undercurrent. Although the 24^o and 26^o isotherm outcrop at roughly the correct position, the stratification is stronger than in the observations and the 20^o isotherm outcrops near the Equator whereas it stays subsurface in the observations. The stratification in the thermocline in the northern and southern flanks is also much stronger than in the observations. Generally, ICON-ESM produces too cold conditions above the thermocline (Figure 17) which is responsible for a pronounced cold bias along the Equator in the Pacific and possibly affects ENSO variability (see section 4.2.7). The section along the Equator confirms that the position of the EUC's core depth is well captured in the model, as are the amplitudes of the EUC and the westward flow near the surface. The weaker EUC seen in (b) is more related to a biased longitudinal positioning of the maximum strength of the EUC.

4.2.6 Sea ice

Sea-ice thickness distributions and extents for late winter and late summer are presented in Figures 20 and 21 for both hemispheres. Arctic and Antarctic sea-ice edges are reproduced in good agreement with the EUMETSAT OSI SAF data set (EUMETSAT, 2015) for the respective winter seasons but summer ice cover is clearly underestimated. In particular in the Arctic, simulated summer sea-ice concentrations are above 15% only between Fram Strait and the North Pole whereas observational data show sea-ice extents almost reaching the Canadian and Siberian coasts. Summer sea ice is also very thin, rarely reaching 1 m thickness. Winter sea-ice thickness is overly too homogeneous and lacks the typical pile-up of thicker ice (exceeding 3 m thickness) that is seen in the PIOMAS Arctic sea-ice volume reanalysis (Schweiger et al., 2011) (compare also: `http://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/piomas-monthly-thickness-map`.) Deficits in the sea-ice thickness distributions are likely related to biases in the atmospheric circulation as indicated by the sea-level pressure errors in Figure 6. Mismatches in the seasonal amplitude and the too thin summer ice need further attention. During the tuning process we tried different options for albedo parameters or the lead-closing

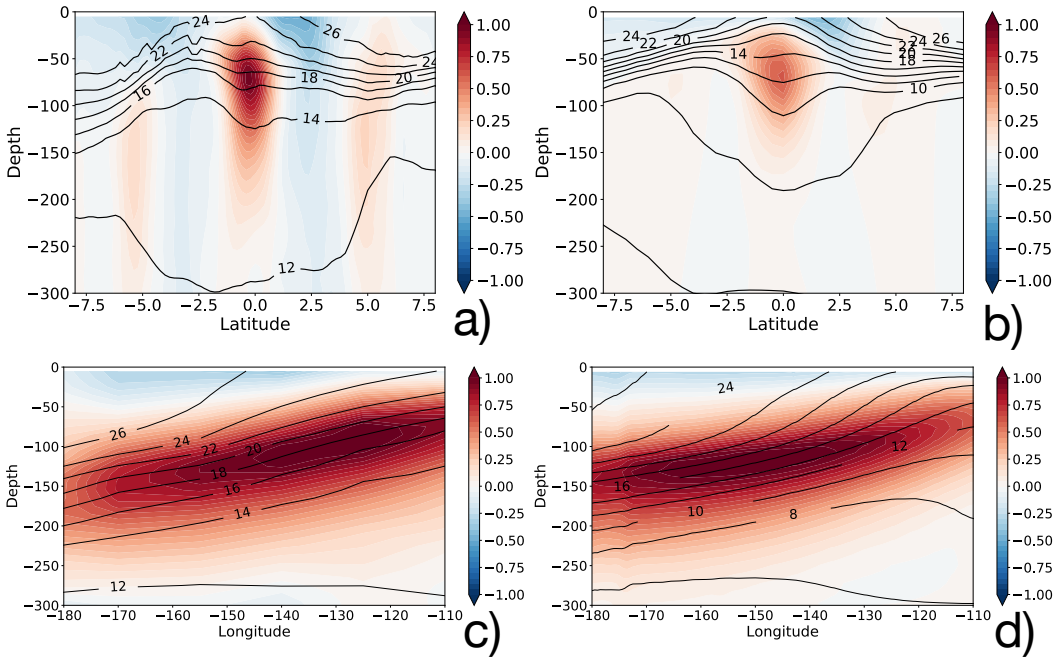

Figure 19. Tropical Pacific zonal velocities (color shading) and temperatures (contour lines) in the ICON-ESM simulation (left) compared to the observation-based estimate by Johnson et al. (2002) (right) at 110^o W (upper row) and along the Equator (lower row).

scheme. These were either unsuccessful in improving summer sea ice or lead to overly large sea-ice extent in winter (in particular in the Labrador Sea) with negative consequences for deep water formation and the strength of the AMOC.

4.2.7 Variability simulated by ICON-ESM

The El-Nino-Southern Oscillation (ENSO) is one of the most important modes of tropical variability with global teleconnections. To analyse characteristic features of ENSO and related quantities we apply the Climate Variability Diagnostics Package (CVDP; Phillips et al., 2014). CVDP allows assessment of simulations in comparison with observation to be carried out in a consistent way. CVDP results presented here can easily be compared with collections of CMIP5 and CMIP6 model evaluations carried out by NCAR (`www.cesm.ucar.edu/workinggroups/CVC/cvdp/data-repository.html`). CVDP provides also links to observational data sets for comparison (ERSST (Smith et al., 2008) and HADSST (Rayner et al., 2003) for sea surface temperature and GPCP (Adler et al., 2018) for precipitation). We show here a subset of ENSO features that are based on SST time series from the Nino3.4 region. First, we present spectra (Fig. 22) obtained for boreal winter (DJF) conditions from the five-member historical ensemble (red lines) and from the piControl simulation (blue lines). The latter was split into three sections of 160 years. All time series show a peak near three years and similar variability in the control and historical runs. The spectra are more pronounced than those from the CMIP6 version of the MPI-ESM-LR and the level of variability is considerably higher than the observed spectra. ERSST data as well for the entire period (ERSST5, 1920-2014) as well as those from the last 35 years (ERSST5_1) show a much broader spectrum and indicate less strong variability on the three-year scale. Regarding the level of spectral power, ICON-ESM is not an exception compared to other CMIP5 and CMIP6 models according to the NCAR repository, but it belongs certainly not to the better performing models. We presume

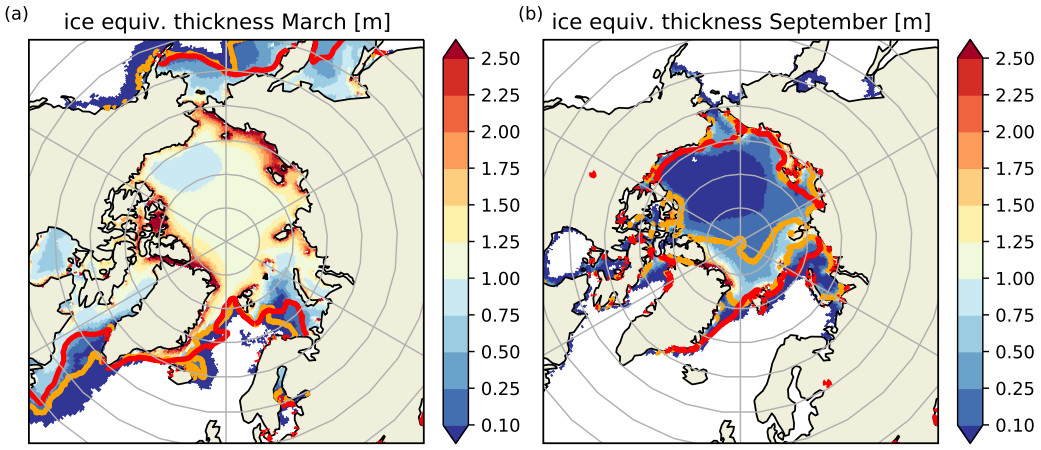

Figure 20. Ensemble mean sea-ice thickness (shading) for a) March and b) September in the northern hemisphere. Thick colored lines depict the 15% sea-ice extent boundary in the simulations (orange) and from the EUMETSAT OSI SAF observational data set (red).

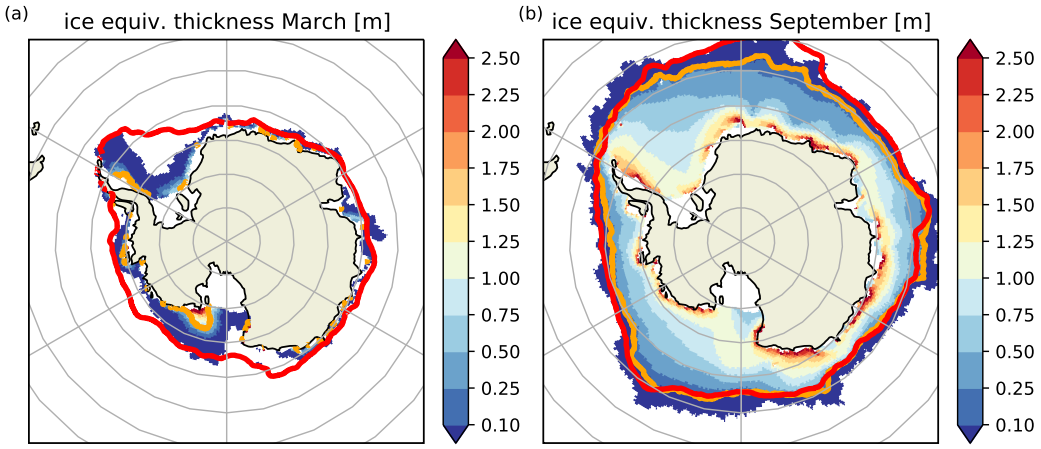

Figure 21. Ensemble mean sea ice thickness (shading) for a) March and b) September in the southern hemisphere. Thick colored lines depict the 15% ice extent boundary in the simulations (orange) and from the EUMETSAT OSI SAF observational data set (red).

that the sharp spectral peak at three years is related to the pronounced cold bias along the equatorial Pacific (Fig. 17) and the related upper ocean stratification. Spatial composites of boreal winter ENSO-related anomalies of temperature and precipitation are shown in Figures 23a and 23b, respectively. While the general patterns are reproduced well, there are deficits in the amplitude of SST variations at the Equator. In particular, positive anomalies reach too far into the warm pool region, where they also shift the positive precipitation anomaly too far to the west (Fig. 23b).

4.3 Ocean biogeochemistry simulated by HAMOCC

The performance of ICON-ESM in simulating the ocean carbon cycle is evaluated in the piControl run. We compare the surface water pCO_2 averaged over the last 10 years of the spin-up run to the piControl results from MPI-ESM 1.2-LR (Mauritsen et al., 2019)

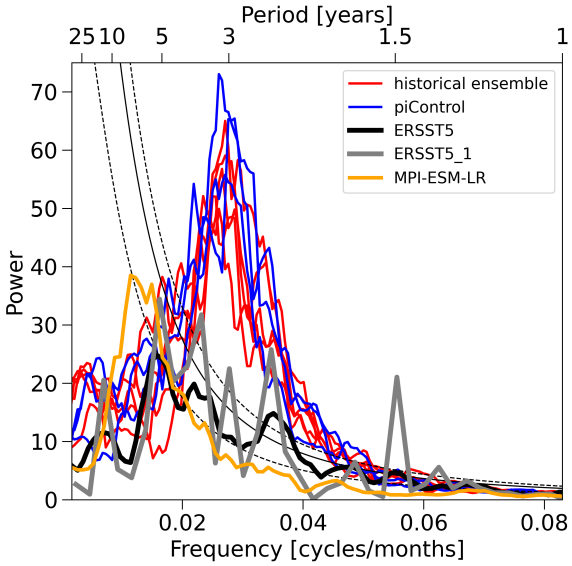

Figure 22. El Nino - Southern Oscillation (ENSO) variability derived from Nino3.4 time series from (red) the historical ensemble, (blue) the piControl simulation, and (orange) a MPI-ESM-LR historical simulation (Mauritsen et al., 2019). The piControl experiment was subdivided into three 160-year long sections. Estimates from the observational ERSST5 data set are included, where ERSST5_1 denotes a subset from the last 35 years). TODO: refrence ERSST!

(Fig. 24). The overall regional patterns are well reproduced in this simulation with maximum values detected in the tropical Pacific and minimum values in the extra-tropical regions; but differences are also detected. ICON-ESM simulates a lower pCO_2 than MPI-ESM 1.2-LR in the equatorial Pacific, most regions of the subtropical Pacific and along the Kuroshio Current. On the contrary, in the Labrador Sea, Barents Sea and along the southern Chilean coast, ICON-ESM simulates higher pCO_2 than MPI-ESM 1.2-LR. The simulated annual global flux of CO_2 into the ocean is about 0.05 PgC/yr, representative for the assumption of the pre-industrial steady-state condition.

To understand the drivers for the difference in the surface water pCO_2 in the two models, we decompose the changes in pCO_2 to changes in dissolved inorganic carbon (DIC), total alkalinity (TA), SST and SSS following Menviel et al. (2018). The sum of changes in DIC and TA are referred to as biogeochemical contributions and the sum of changes in SST and SSS are the physical contributions, respectively. The change in pCO_2 due to these contributions is calculated in each model and the difference is illustrated (Fig. 24c,d). The sum of the biogeochemical and physical contributions (not shown) closely matches the difference between ICON-ESM and MPI-ESM pCO_2 shown in Fig. 24b. The lower pCO_2 in ICON-ESM in the eastern equatorial Pacific can be attributed to the lower SST, while in the subtropical Pacific, western equatorial Pacific and along the Kuroshio Current it is due to the biogeochemical contribution, mostly counteracting the impact of the circulation. Along the Labrador Current, the higher pCO_2 in ICON-ESM is attributed to biogeochemical changes, a result of lower simulated TA in ICON-ESM. In the Barents Sea, the higher pCO_2 results from the combined effect of a higher SST and higher DIC simulated by ICON-ESM.

These HAMOCC results represent a first-order tuning attempt and further simulations will be required to obtain a fully tuned HAMOCC model. Therefore, the evaluation of ocean biogeochemistry in a historical run will be the topic of a future study.

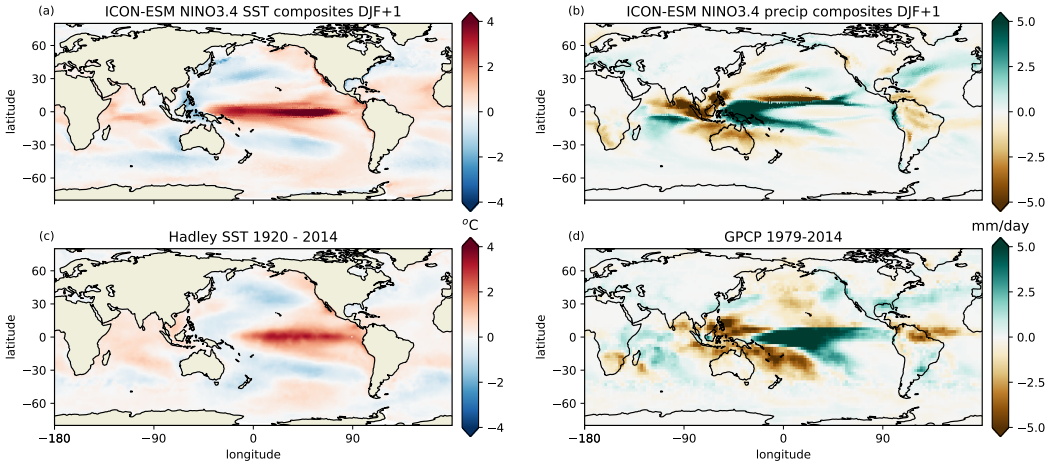

Figure 23. Spatial composites of ENSO (Nino3.4) events for a), c) SST, and b), d) precipitation from a), b) the historical ensemble mean, c) the HadSST, and d) the GPCP precipitation data sets.

4.4 Idealized CO_2 increase simulations and climate sensitivity

Two measures of the model's response to changes in radiative forcing are computed using the idealized CO_2-increase experiments in CMIP6 DECK: Transient Climate Response (TCR) and Equilibrium Climate Sensitivity (ECS). TCR is computed from the experiment where CO_2 increases by 1 percent per year (1pctCO2). The TCR is estimated as the global temperature (TAS) increase around the time of CO_2-doubling, which happens after 70 years of simulation. To minimize effects of internal variability we take a 20-year average between years 60 and 79 as suggested by Meehl et al. (2020) and we arrive at a TCR of 2.1K. This value is slightly higher than the 1.8 K derived for the CMIP6 version of MPI-ESM-LR and in the middle of the range found in the CMIP6 multi-model assessment by Meehl et al. (2020). Differences in TCR are often related to variations in the ocean heat content changes under global warming, which could be different in MPIOM and ICON-O due to different formulations of vertical mixing and the GM parameters. ECS is estimated using the DECK "abrupt4xCO2" experiment applying the method of Gregory et al. (2004), which has also been used by Meehl et al. (2020). We perform a linear regression between the global mean TAS and the net downward radiative flux at the top-of-atmophere over 150 years of simulation. ECS is then estimated from an extrapolation of the regression line to zero net heat imbalance (Fig. 25b). This procedure results in an ECS of 3.7 K, which can be compared with the value of 2.9 in MPI-ESM-LR. MPI-ESM and ICON-ESM are quite similar in their estimate of the effective forcing (estimated as the crossing of the regression lines with the y-axis in Fig. 25b), but the slopes of the regression lines are considerably different. We note that the change of temperature over time differs for the later part of the experiments. Redoing the ECS estimation excluding the first 20 years in the regression, we find a much higher value of 4.3 K in ICON, whereas there is only a relatively small change to 3.1 K in MPI-ESM. This difference is likely related to different evolution of local or regional feedbacks (Armour et al., 2013). The higher TCR and ECS in ICON-ESM compared to MPI-ESM-LR may be responsible for part of the more pronounced warming in the second half of the 20^{th} century (Fig. 4). However, for the ICON-ESM no effort was made to tune the climate sensitivity in order to better match the historical record as it was done for MPI-ESM1.2 (Mauritsen & Roeckner, 2020).

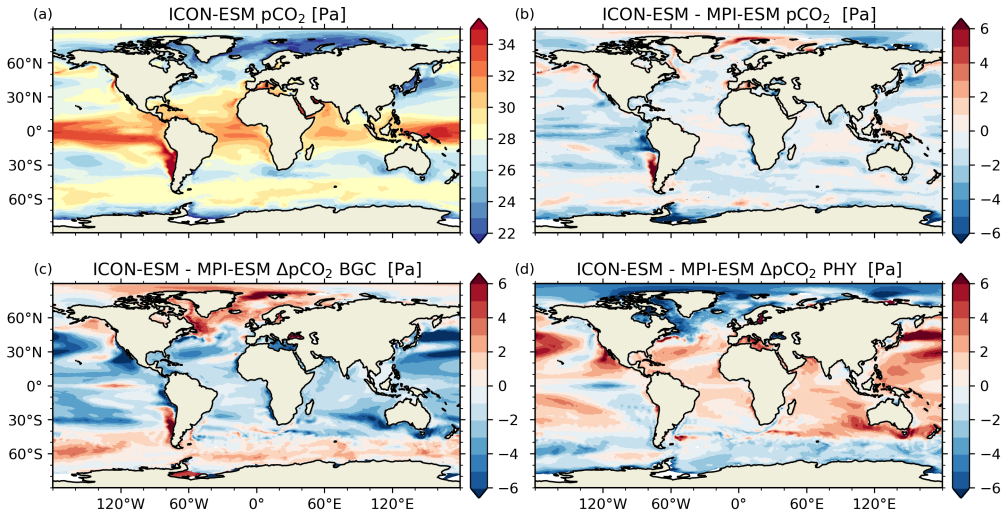

Figure 24. Surface pCO_2 in a) ICON-ESM piControl simulations and b) ICON-ESM minus MPI-ESM piControl simulations. The contribution of biogeochemical factors (alkalinity and DIC) (c) and physical factors (SST and SSS) (d) to the difference in ICON-ESM and MPI-ESM pCO_2 are also shown.

5 Discussion

During the tuning phase we have explored different parameter settings and, in the case of ocean vertical mixing, the choice between two different parameterization schemes (TKE, PP, see section 2.2). We started the coupled experiments using parameter settings inherited from the stand-alone ocean and atmosphere simulations. While these gave good or at least acceptable results in ICON-A (Crueger et al., 2018) and ICON-O (Korn et al., manuscript in preparation), solutions meeting our tuning goals (see Section 3) were much harder to obtain.

Problematic biases increased or became more apparent as feedbacks between the coupled components evolved. For example, the ICON-A AMIP simulation described in Crueger et al. (2018) exhibits relatively good skill scores, but features strong positive sea-level pressure biases in high northern latitudes (their Figure 3). In the coupled simulation, this error is accompanied by too weak winds over the subpolar North Atlantic that could lead to biases in ocean circulation, water mass transformation and, eventually, to a strong reduction or collapse of the AMOC. Changing parameter settings in the SSO parameterization (see Table 1) turned out to be an effective way to reduce the SLP bias. However, for instance, small values of *gkdrag* reducing SLP biases also led to stronger errors in upper tropospheric and stratospheric zonal wind strength and distribution. Therefore, the results presented here are the results of compromises. Avoiding detrimental effects in some key quantities, such as the collapse of the AMOC or a freeze-over of the Labrador Sea in the ocean, required sometimes parameter settings in the atmosphere that turned out to be sub-optimal in terms of atmospheric performance skill scores (Figure 5).

In general, we found that tuning choices had often complex and unexpected effects in the coupled system. Partly, these were hard to grasp as they required relatively long adjustment times. Moreover, obtaining a tuning target is often dependent on several parameter settings and these may influence each other. We demonstrate this with the ex-

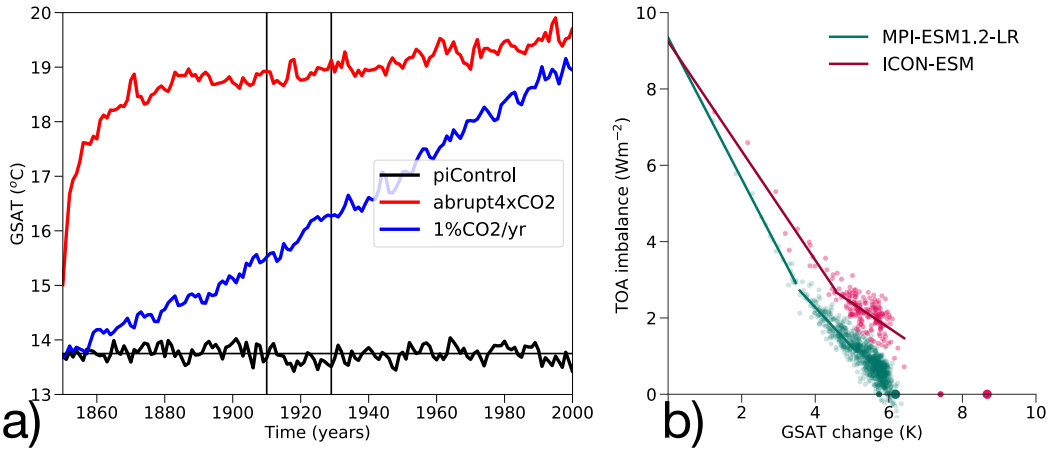

Figure 25. Estimating the Transient Climate Response (TCR) and Equilibrium Climate Sensitivity (ECS) from idealized DECK experiments. a) Evolution of global surface air temperature in the $1pctCO2$ (blue) and $abrupt4XCO2$ (red) together with the $piControl$ experiment (black). Vertical lines indicate the period for the TCR calculation, b) Gregory-style regression analyses to estimate ECS in ICON-ESM (dark red) and in the CMIP6 model MPI-ESM1.2.

ample of the AMOC strength taken from several experiments with different parameter settings. ICON-O stand-alone experiments, where only the coefficients for isoneutral diffusion K and eddy-induced diffusion κ have been changed, indicated that larger K and κ lead to weaker overall AMOC strength, likely due to a stronger flattening of isopycnals (not shown). However, an inspection of the multitude of experiments leads us to conclude that it is possible to arrive at any AMOC state even for K and *kappa* set to zero. We found fewer solutions with weak overturning at high K and *kappa*, but this may be due to the smaller number of experiments. It is also difficult to relate the AMOC strength uniquely to other parameters. For example, the experiments run with $K/kappa$ equal 400 ms^{-2} in Figure 26 come with various settings of the wave drag parameter *gkdrag* and an inspection of this column in isolation would lead us to conclude that lower wave drag parameters produce higher ocean overturning, possibly due to the effect on wind and wind stress discussed above. But again, looking at all experiments using a moderately low *gkdrag* of 0.02 we find an AMOC range between 7.6 and almost 25 Sv.

It is obvious that another round of tuning and an even more extensive set of sensitivity experiments could have improved some of the shortcomings mentioned above. We decided, however, to finalize "version 1" at this stage to set a milestone in the model development. Moreover, model development activities in the component models have continued in parallel and we prefer to postpone a second round of major tuning efforts until some innovations can be included in the standard model configurations. We mention here two developments in ICON-O: the z* vertical coordinate and a newly developed dynamical sea-ice model. z* is designated to replace "z-level" as standard coordinate and comes together with an improved representation of the bottom topography in the form of "partial cells", which was already implemented in MPIOM and its predecessors. The implementation process includes a slight reorganization of level distribution in the upper ocean and we plan to achieve a better representation of the mixed layer processes and mixing by re-assessing the TKE parameterization. Mehlmann and Korn (2021) have developed a novel sea-ice dynamics formulation, which is based on an analogue of the Arakawa-CD grid. The CD-grid placement has appealing resolving properties at high spatial resolution compared to traditionally used discretizations (Mehlmann et al., 2021). Furthermore, the development allows a straightforward coupling to the Arakawa C-grid-

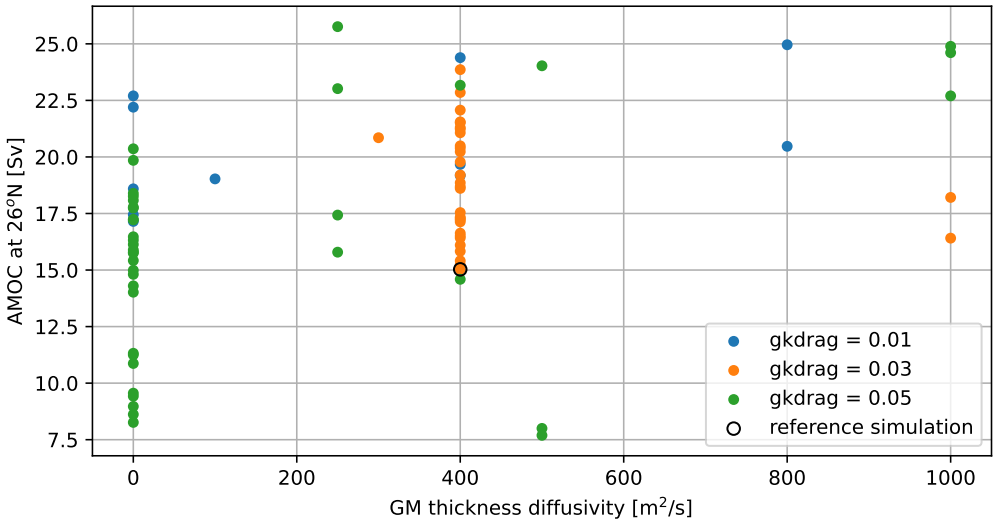

Figure 26. Scatterplot relating the eddy diffusion parameter κ with the strength of the AMOC at 26°N. The black circle indicates the standard experiment (piControl) and colors refer to different choices of the gravity wave drag parameter *gkdrag*.

like discretization used in ICON-O. As the new sea ice dynamics are realized on the same grid as ICON-O the coupling requires no rotations and promises a better representation of the bathymetry. While the deficits in the simulated sea-ice climatology documented here may be related to problems in sea-ice thermodynamics, we expect improvement from the new dynamics, e.g., for the representation of narrow passages and the related ice transports. By affecting fresh-water exchanges, the latter could lead to improvements in water mass properties and air-sea exchanges for example in the Labrador Sea.

Ongoing development work will further explore ICON-specific opportunities like grid refinement in ICON-O (Logemann et al., 2021) and nesting in the atmosphere (Jungandreas et al., 2021). The excellent scaling capabilities of ICON have been documented in the DYAMOND project with ICON-A setups between 80 and 2.5 km (Stevens et al., 2019; Hohenegger et al., 2020) and are further explored in very high resolution coupled setups with grid spacing of a few kilometers in the DYAMOND-WINTER project (`https://www.esiwace.eu/services/dyamond/winter`).

While the ICON-A version described herein uses the physical parameterization package inherited from ECHAM6 (Stevens et al., 2013; Giorgetta et al., 2018), which was designed for "climate" applications at grid sizes from 50 to 350 km, higher-resolution ICON application will require other choices. At km-scale, some parameterizations may become obsolete or will be better represented by schemes from the numerical weather-prediction version of ICON (Zängl et al., 2015). Therefore, a longer-term goal of the ICON community is the development of a system for "seamless" predictions from weather to climate scales.

6 Summary and conclusions

In this paper, we have documented ICON-ESM (V1.0), the first coupled model that is based on the ICON framework (Zängl et al., 2015) with its isosahedral grid concept. We have presented the first simulations with the coupled ICON-ESM (Lorenz et al., 2021) confronting it with the task to deliver reasonable results in a well-defined experimental

framework, the CMIP6 DECK simulations. Coupling of the newly developed component models and the tuning of the coupled model turned out to be more challenging than expected. Tuning choices that revealed robust results in stand-alone simulations needed adjustment or reconfiguration in the presence of coupled feedbacks. The performance of the ICON-ESM reported in this paper can be summarized as follows:

1. We were able to fulfil the primary tuning goals: The 500-year long piControl simulation exhibits little drift in key quantities like GSAT, radiation, sea-ice cover, and the AMOC TOA radiation is well balanced with little remaining drift. GSAT and other global quantities are close to observational estimates for the pre-industrial climate.

2. A set of five historical simulations reproduces the observed global surface temperature evolution largely in agreement with observational data sets, albeit with somewhat too strong warming in the second half of the 20th century.

3. ICON-ESM's climate sensitivities (TCR, ECS) are slightly higher than the corresponding values from MPI-ESM, but well in the range of the CMIP6 multi-model ensemble.

4. The present-day climate simulated for the last decades reproduces largely the climatology from observations and reanalyses, but biases are often larger than in ICON-ESM's predecessor model MPI-ESM1.2 and other CMIP6 models.

Problematic issues identified in the analyses are: In the atmosphere, the representation of both low-level and high-level clouds showed considerable deficits. Bias patterns, which were already identified in stand-alone atmosphere simulations became more pronounced in the coupled system. Counter-tuning that could have improved cloud distributions had negative side effects on other parts of the climate system. Skill scores based on wind and temperature data are worse than in MPI-ESM. In particular zonal wind biases remain and there is some indication that the parameterization of sub-grid scale orographic effects requires further attention.

Overall, the rather zonal pattern of the LST biases hints to the global atmospheric circulation as their major cause. Nevertheless, the land surface temperatures show regional biases that may be traced to a misrepresentation of surface albedo values. The biases over Antarctica and Greenland partly reflect albedo deviations over glaciers. Errors in land temperatures over Eurasia are substantially caused by a too low snow cover and the associated snow-albedo feedback. The biases could be partly improved by changes in the near-infrared and visible albedo settings in JSBACH 4, but others are complex and related to deficits in snow cover or soil and canopy albedo biases.

In the ocean, SST and SSS climatologies show bias patterns and magnitudes comparable to other coarse-resolution CMIP5 or CMIP6 models, whereas the relatively large sub-surface biases are of concern in particular in the tropical oceans. Here, a re-tuning of the vertical mixing scheme might alleviate misrepresentation of wind-induced mixing. From improvements in the upper-ocean stratification we expect also positive effects on ENSO variability and processes in the tropical oceans. Northern hemisphere sea ice concentrations is somewhat too high in winter with too much ice cover, for example in the Labrador Sea. On the other hand, the seasonal cycle is too strong leading to too thin sea-ice in summer with only half of the Arctic basis being ice-covered in summer. The seasonal cycle is also too strong in the southern hemisphere.

ENSO variability in ICON-ESM is less close to observations than MPI-ESM's, but similar to many other CMIP5 and CMIP6 models. The spectra derived from Nino3.4 time-series of the historical simulations show a too narrow peak at about three years and an overly high level of variance. The spatial patterns associated to ENSO variability agree with the observed ones, but there are too strong signals in the warm pool and in the Indian Ocean.

We emphasize that the present configuration is the first milestone in establishing ICON-ESM as a highly flexible modelling system. ICON-ESM V1 serves as a basis for further developments that will take advantage of ICON-specific properties such as spatially varying resolution, and coupled configurations at resolutions of a few kilometers in atmosphere, land, and ocean.

7 Data Availability Statement

The data from the ICON-ESM V1.0 DECK simulations are available at the CMIP6 repository of the Earth System Grid Federation (Lorenz et al., 2021) and can be accessed at `http://esgf-data.dkrz.de/search/cmip6-dkrz/`. The model code of ICON is available to individuals under licenses (`https://mpimet.mpg.de/en/science/modeling-with-icon/code-availability`). By downloading the ICON source code, the user accepts the licence agreement. The source code of the ICON-ESM-V1.0 used in this study, primary data, and sripts used in the analyses and for producing the figures can be obtained from the WDCC Long Term Archive (`http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=RUBY-0_ICON-_ESM_V1.0_Model`).

Acknowledgments

We thank the German Climate Computer Centre DKRZ for providing the computational resources. TI was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 101003536 (ESM2025 – Earth System Models for the Future) and under grant agreement No 820989 (COMFORT)". TI and FC were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2037 'Climate Climatic Change and Society' (CLICCS) - Project Number: 390683824 and by the European Union's Horizon 2020 research and innovation programme under grant agreement number 773421 - project "Nunataryuk"." This paper is a contribution to the project S2 (PIs JJ and NB) of the Collaborative Research Centre TRR 181 "Energy Transfers in Atmosphere and Ocean" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project nr. 274762653. HP received funding by DWD's "Innovation Programme for Applied Researches and Developments" (IAFE VH3.5). FZ was supported by the German Federal Ministry of Education and Research (BMBF) as a Research for Sustainability initiative (FONA) through the project PalMod (FKZ: 01LP1502A) and by the European Union's Horizon 2020 research and innovation programme under grant agreement No 823988 (ESiWACE2). The CMIP6 participation received funding by the BMBF via the project CMIP6-DICAD (FKZ: 01LP1605A).

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., ... others (2003). The version-2 global precipitation climatology project (gpcp) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*(6), 1147–1167.

Adler, R. F., Sapiano, M. R., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., ... others (2018). The global precipitation climatology project (gpcp) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4), 138.

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, *26*(13), 4518–4534.

Becker, J., Sandwell, D., Smith, W., Braud, J., Binder, B., Depner, J., ... others (2009). Global bathymetry and elevation data at 30 arc seconds resolution: Srtm30_plus. *Marine Geodesy*, *32*(4), 355–371.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., ...

others (2013). Multiple stressors of ocean ecosystems in the 21st century: projections with cmip5 models. *Biogeosciences*, *10*(10), 6225–6245.

Borchert, S., Zhou, G., Baldauf, M., Schmidt, H., Zängl, G., & Reinert, D. (2019). The upper-atmosphere extension of the icon general circulation model (version: ua-icon-1.0). *Geoscientific Model Development*, *12*(8), 3541–3569.

Brovkin, V., Raddatz, T., Reick, C. H., Claussen, M., & Gayler, V. (2009). Global biogeophysical interactions between forest and climate. *Geophysical research letters*, *36*(7). doi: 10.1029/2009GL037543

Cescatti, A., Marcolla, B., Vannan, S. K. S., Pan, J. Y., Román, M. O., Yang, X., ... others (2012). Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network. *Remote sensing of environment*, *121*, 323–334.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J., Minnis, P., ... Zeng, S. (2010). The gcm-oriented calipso cloud product (calipso-goccp). *Journal of Geophysical Research: Atmospheres*, *115*(D4).

Crueger, T., Giorgetta, M. A., Brokopf, R., Esch, M., Fiedler, S., Hohenegger, C., ... Stevens, B. (2018). ICON-A, The Atmosphere Component of the ICON Earth System Model: II. Model Evaluation [Journal Article]. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1638-1662. doi: 10.1029/2017ms001233

Cunningham, S. A., Alderson, S. G., King, B. A., & Brandon, M. A. (2003). Transport and variability of the Antarctic circumpolar current in Drake Passage. *J. Geophys. Res., 108 (C5)*, 8084. doi: 10.1029/2001JC001147

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., ... others (2020). The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2).

Danilov, S., Wang, Q., Timmermann, R., Iakovlev, N., Sidorenko, D., Kimmritz, M., ... Schröter, J. (2016). Finite-Element Sea Ice Model (FESIM), version 2. *Geoscientific Model Development Discussions*, *8*, 1747–1761.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., ... others (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

de Vrese, P., Stacke, T., Kleinen, T., & Brovkin, V. (2021). Diverging responses of high-latitude co 2 and ch 4 emissions in idealized climate change scenarios. *The Cryosphere*, *15*(2), 1097–1130. doi: 10.5194/tc-15-1097-2021

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

Drews, A., Greatbatch, R. J., Ding, H., Latif, M., & Park, W. (2015). The use of a flow field correction technique for alleviating the North Atlantic cold bias with application to the Kiel Climate Model [Journal Article]. *Ocean Dynamics*, *65*(8), 1079-1093. doi: 10.1007/s10236-015-0853-7

Ekici, A., Beer, C., Hagemann, S., Boike, J., Langer, M., & Hauck, C. (2014, apr). Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model. *Geoscientific Model Development*, *7*(2), 631–647. doi: 10.5194/gmd-7-631-2014

EUMETSAT, O. (2015). *Global sea ice concentration reprocessing dataset 1978–2015 (v1. 2).*

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

Fieg, K., Gerdes, R., Fahrbach, E., Beszczynska-Möller, A., & Schauer, U. (2010).

Simulation of oceanic volume transports through Fram Strait 1995–2005. *Ocean Dyn.*, *60*, 491–502. doi: 10.1007/s10236-010-0263-9

Ganachaud, A., & Wunsch, C. (2003). Large-scale ocean heat and freshwater transports during the World Ocean Circulation Experiment [Journal Article]. *Journal of Climate*, *16*(4), 696-705. doi: Doi10.1175/1520-0442(2003)016⟨0696:Lsohaf⟩2.0.Co;2

Gaspar, P., Grégoris, Y., & Levefre, J.-M. (1990). A Simple Eddy Kinetic Energy Model for Simulations of the Oceanic Vertical Mixing Tests at Station Papa and Long-Term Upper Ocean Study Site. *Journal of Geophysical Research: Oceans*, *95*(C9), 16179-16193.

Gent, P., & McWilliams, J. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, *20*, 150–155.

Gent, P. R., Large, W. G., & Bryan, F. O. (2001). What sets the mean transport through Drake Passage? *Journal of Geophysical Research: Oceans*, *106*(C2), 2693–2712.

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., ... Stevens, B. (2018). ICON-A, the Atmosphere Component of the ICON Earth System Model: I. Model Description. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1613-1637. doi: 10.1029/2017ms001242

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Bottinger, M., ... Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597. doi: 10.1002/jame.20038

Goll, D. S., Winkler, A. J., Raddatz, T., Dong, N., Prentice, I. C., Ciais, P., & Brovkin, V. (2017). Carbon–nitrogen interactions in idealized simulations with jsbach (version 3.10). *Geoscientific Model Development*, *10*(5), 2009–2030. doi: 10.5194/gmd-10-2009-2017

Gordon, A. L., Sprinthall, J., Van Aken, H. M., Susanto, D., Wijffels, S., Molcard, R., ... Wirasantosa, S. (2010). The Indonesian throughflow during 2004–2006 as observed by the INSTANT program. *Dyn. Atmos. Oceans*, *50*, 115–128. doi: 10.1016/j.dynatmoce.2009.12.002

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., ... Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical research letters*, *31*(3), L03205. doi: 10.1029/2003GL018747

Griffies, S. (1998). The gent-mcwilliams skew flux. *Journal of Physical Oceanography*, *28*, 831–841.

Griffies, S., Gnanadesikan, A., Pacanowski, R. C., Larichev, V., Dukowicz, J., & Smith, R. (1998). Isoneutral diffusion in a z-coordinate model. *Journal of Physical Oceanography*, *28*, 831–841.

Gutjahr, O., Brüggemann, N., Haak, H., Jungclaus, J. H., Putrasahan, D. A., Lohmann, K., & von Storch, J.-S. (2021). Comparison of ocean vertical mixing schemes in the Max Plank Institute Earth System Model (MPI-ESM1.2). *Geoscientific Model Development*, *14*, 2317—2349. doi: 10.5194/gmd-14-2317-2021

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J. S., Brüggemann, N., ... Stössel, A. (2019). Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, *12*(7), 3241–3281. doi: 10.5194/gmd-12-3241-2019

Hagemann, S., & Dümenil, L. (1997). A parametrization of the lateral waterflow for the global scale. *Climate Dynamics*, *14*, 17–31. doi: 10.1007/s003820050205

Hanke, M., Redler, R., Holfeld, T., & Yastremsky, M. (2016). YAC 1.2.0: new aspects for coupling software in Earth system modelling. *Geoscientific Model De-*

velopment, *9*(8), 2755–2769. doi: 10.5194/gmd-9-2755-2016

Hansen, B., Østerhus, S., Turrell, W. R., Jónsson, S., Valdimarsson, H., Hátún, H., & Olsen, S. M. (2008). The Inflow of Atlantic Water, Heat, and Salt to the Nordic Seas Across the Greenland–Scotland Ridge. In R. R. Dickson, J. Meincke, & P. Rhines (Eds.), *Arctic–Subarctic Ocean Fluxes: Defining the Role of the Northern Seas in Climate* (pp. 15–43). Dordrecht: Springer Netherlands. doi: 10.1007/978-1-4020-6774-7_2

Hawkins, E., & Sutton, R. (2016). Connecting Climate Model Projections of Global Temperature Change with the Real World. *Bulletin of the American Meteorological Society*, *97*(6), 963–980. doi: 10.1175/Bams-D-14-00154.1

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., ... Zadeh, N. (2019). Structure and Performance of GFDL's CM4.0 Climate Model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. doi: 10.1029/2019ms001829

Hines, C. O. (1997). Doppler-spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 1: Basic formulation. *J. Atm. Sol. Terr. Phys.*, *59*, 371-386.

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., ... Stevens, B. (2020). Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *Journal of the Meteorological Society of Japan. Ser. II*.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., ... Zhang, X. (2019). *input4mips.cmip6.cmip.uofmd.uofmd-landstate-high-2-1-h.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/input4MIPs.11261` doi: 10.22033/ESGF/input4MIPs.11261

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model hamocc: Model architecture and performance as component of the mpi-earth system model in different cmip5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, *5*(2), 287–315.

Jochumsen, K., Quadfasel, D., Valdimarsson, H., & Jónsson, S. (2012). Variability of the Denmark Strait overflow: Moored time series from 1996–2011. *J. Geophys. Res.*, *117*, C12003. doi: 10.1029/2012JC008244

Johnson, G. C., Sloyan, B. M., Kessler, W. S., & McTaggart, K. E. (2002). Direct measurements of upper ocean currents and water properties across the tropical Pacific during the 1990s. *Progress in Oceanography*, *52*(1), 31–61.

Jungandreas, L., Hohenegger, C., & Claussen, M. (2021). Influence of the representation of convection on the mid-holocene west african monsoon. *Climate of the Past*, *17*, 1665–1684. doi: 10.5194/cp-17-1665-2021

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., ... von Storch, J. S. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446. doi: 10.1002/jame.20023

Keeley, S. P. E., Sutton, R. T., & Shaffrey, L. C. (2012). The impact of North Atlantic sea surface temperature errors on the simulation of North Atlantic European region climate. *Quarterly Journal of the Royal Meteorological Society*, *138*(668), 1774–1783. doi: 10.1002/qj.1912

Klocke, D., Brueck, M., Hohenegger, C., & Stevens, B. (2017). Rediscovery of the doldrums in storm-resolving simulations over the tropical atlantic. *Nature Geoscience*, *10*(12), 891–896.

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

Korn, P. (2018). A Structure Preserving Discretization of Ocean Parametrizations

on Unstructured Grids. *Ocean Modelling*, *132*, 73–90.

Korn, P., & Linardakis, L. (2018). A conservative discretization of the shallow-water equations on triangular grids. *Journal of Computational Physics*, *375*, 871–890.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., ... others (2020). Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from cmip6 model projections. *Biogeosciences*, *17*(13), 3439–3470.

Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of geophysics*, *32*(4), 363–403.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research-Atmospheres*, *124*(12), 6307–6326. doi: 10.1029/2018jd029522

Liu, X., & Schumaker, L. L. (1996). Hybrid bézier patches on sphere-like surfaces. *Journal of Computational and Applied Mathematics*, *73*(12), 157-172. doi: 10.1016/0377-0427(96)00041-6

Logemann, K., Linardakis, L., Korn, P., & Schrum, C. (2021). Global tide simulations with icon-o: testing the model performance on highly irregular meshes. *Ocean Dynamics*, *71*(1), 43–57.

Lohmann, U., & Roeckner, E. (1996). Design and performance of a new cloud microphysics scheme developed for the echam general circulation model. *Climate Dynamics*, *12*(8), 557–572.

Lorenz, S., Jungclaus, J., Schmidt, H., Haak, H., Reick, C., Schupfner, M., ... Ilyina, T. (2021). *Mpi-m icon-esm-lr model output prepared for cmip6 cmip 1pctco2.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.6433` doi: 10.22033/ESGF/CMIP6.6433

Lott, F. (1999). Alleviation of stationary biases in a gcm through a mountain drag parameterization scheme and a simple representation of mountain lift forces. *Monthly weather review*, *127*(5), 788–801.

Maerz, J., Six, K. D., Stemmler, I., Ahmerkamp, S., & Ilyina, T. (2020). Microstructure and composition of marine aggregates as co-determinants for vertical particulate organic carbon transfer in the global ocean. *Biogeosciences*, *17*(7), 1765–1803.

Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., & Roske, F. (2003). The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modelling*, *5*(2), 91–127. doi: 10.1016/S1463-5003(02)00015-X

Martin, J. H., Knauer, G. A., Karl, D. M., & Broenkow, W. W. (1987). Vertex: carbon cycling in the northeast pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, *34*(2), 267–285.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. doi: 10.1029/2018ms001400

Mauritsen, T., & Roeckner, E. (2020). Tuning the MPI-ESM1.2 Global Climate Model to Improve the Match With Instrumental Record Warming by Lowering Its Climate Sensitivity. *Journal of Advances in Modeling Earth Systems*, *12*(5), e2019MS002037. doi: 10.1029/2019MS002037

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., ... Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*, M00a01. doi: 10.1029/2012ms000154

Mauritsen, T., Svensson, G., Zilitinkevich, S. S., Esau, I., Enger, L., & Grisogono, B. (2007). A total turbulent energy closure model for neutrally and stably strat-

ified atmospheric boundary layers. *Journal of Atmospheric Sciences*, *64*(11), 4113–4126.

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

Mehlmann, C., Danilov, S., Losch, M., Lemieux, J.-F., Hutter, N., Richter, T., ... Korn, P. (2021). Simulating linear kinematic features in viscous-plastic sea ice models on quadrilateral and triangular grids. *http://arxiv.org/abs/2103.04431*.

Mehlmann, C., & Korn, P. (2021). Sea-ice dynamics on triangular grids. *Journal of Computational Physics*, *428*, 110086.

Menviel, L., Spence, P., Yu, J., Chamberlain, M., Matear, R., Meissner, K., & England, M. H. (2018). Southern hemisphere westerlies as a driver of the early deglacial atmospheric co2 rise. *Nature Communications*, *9*(1), 1–12.

Milinski, S., Bader, J., Haak, H., Siongco, A. C., & Jungclaus, J. H. (2016). High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic. *Geophysical Research Letters*, *43*(19), 10455–10462. doi: 10.1002/2016gl070530

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research-Atmospheres*, *117*, D08101. doi: 10.1029/2011jd017187

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... others (2018a). A higher-resolution version of the max planck institute earth system model (mpi-esm1. 2-hr). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413.

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... Marotzke, J. (2018b). A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413. doi: 10.1029/2017ms001217

Nabel, J. E., Naudts, K., & Pongratz, J. (2020, jan). Accounting for forest age in the tile-based dynamic global vegetation model JSBACH4 (4.20p7; git feature/forests) - a land surface model for the ICON-ESM. *Geoscientific Model Development*, *13*(1), 185–200. doi: 10.5194/gmd-13-185-2020

Nordeng, T.-E. (1994, 09). *Extended versions of the convective parametrization scheme at ecmwf and their impact on the mean and transient activity of the model in the tropics* (No. 206). Shinfield Park, Reading: ECMWF. Retrieved from `https://www.ecmwf.int/node/11393` doi: 10.21957/e34xwhysw

Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H., & Marotzke, J. (2013). Arctic sea-ice evolution as modeled by max planck institute for meteorology's earth system model. *Journal of Advances in Modeling Earth Systems*, *5*(2), 173–194.

Nowlin Jr., W. D., & Klinck, J. M. (1986). The physics of the Antarctic Circumpolar Current. *Rev. Geophys.*, *24(3)*, 469–491. doi: 10.1029/RG024i003p00469

Pacanowski, R. C., & Philander, S. G. H. (1981). Parameterization of vertical mixing in numerical models of tropical oceans. *Journal of Physical Oceanography*, *11*(11), 1443–1451. doi: 10.1175/1520-0485(1981)011⟨1443:POVMIN⟩2.0.CO;2

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model hamocc. *Journal of Advances in Modeling Earth Systems*, *9*(1), 438–464.

Phillips, A. S., Deser, C., & Fasullo, J. (2014). Evaluating modes of variability in climate models. *Eos, Transactions American Geophysical Union*, *95*(49), 453–

455. doi: 10.1002/2014EO490002

Pincus, R., & Stevens, B. (2013). Paths to accuracy for radiation parameterizations in atmospheric models. *Journal of Advances in Modeling Earth Systems*, *5*(2), 225–233.

Pongratz, J., Reick, C., Raddatz, T., & Claussen, M. (2008). A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochemical Cycles*, *22*(3). doi: 10.1029/2007GB003153

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., ... Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(D14).

Redi, M. (1982). Oceanic Isopycnal Mixing by Coordinate Rotation [Journal Article]. *Journal of Physical Oceanography*, *12*, 1154–1158.

Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3), 303–312.

Reick, C. H., Gayler, V., Goll, D., Hagemann, S., Heidkamp, M., Nabel, J. E. M. S., ... Wilkenskjeld, S. (2021). JSBACH 3 - The land component of the MPI Earth System Model: Documentation of version 3.2. *Berichte zur Erdsystemforschung*, *240*. doi: 10.17617/2.3279802

Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*(3), 459–482. doi: 10.1002/jame.20022

Riddick, T. (2021). *Generation of hd parameters files for icon grids: technical note.* doi: 10.17617/2.3336390

Rieger, D., Bangert, M., Bischoff-Gauss, I., Förstner, J., Lundgren, K., Reinert, D., ... others (2015). Icon–art 1.0–a new online-coupled model system from the global to regional scale. *Geoscientific Model Development*, *8*(6), 1659–1676.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., ... Tompkins, A. (2003). *The atmospheric general circulation model ECHAM 5. PART I: Model description* (Vol. 349). Hamburg: Max-Planck-Institut für Meteorologie. Retrieved from `https://pure.mpg.de/pubman/item/item_995269_6/component/file_3192562/349-Report-txt.pdf` doi: 10.17617/2.995269

Rossby, T., & Flagg, C. (2012). Direct measurement of volume flux in the Faroe-Shetland Channel and over the Iceland-Faroe Ridge. *Geophys. Res. Lett.*, *39*, L07602. doi: 10.1029/2012GL051269

Schaaf, C., & Wang, Z. (2015). MCD43C3 MODIS/Terra+ Aqua BRDF/Albedo Albedo Daily L3 Global 0.05 Deg CMG V006. *NASA EOSDIS Land Processes DAAC. doi*, *10*. doi: doi.org/10.5067/MODIS/MCD43C3.006

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., ... others (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote sensing of Environment*, *83*(1-2), 135–148. doi: 10.1016/S0034-4257(02)00091-3

Schweiger, A., Lindsay, R., Zhang, J. L., Steele, M., Stern, H., & Kwok, R. (2011). Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research-Oceans*, *116*, C00d06. doi: 10.1029/2011jc007084

Séférian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., ... others (2020). Tracking improvement in simulated marine biogeochemistry between cmip5 and cmip6. *Current Climate Change Reports*, 1–25.

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

Six, K. D., & Maier-Reimer, E. (1996). Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles*, *10*(4), 559–583.

Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams,

E., ... McCarthy, G. D. (2018). The North Atlantic Ocean Is in a State of Reduced Overturning. *Geophysical Research Letters*, *45*(3), 1527–1533. doi: 10.1002/2017gl076350

Smith, T. M., Reynolds, R. W., Peterson, T. C., & Lawrimore, J. (2008). Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *Journal of Climate*, *21*(10), 2283–2296. doi: 10.1175/2007JCLI2100.1

Steele, M., Morley, R., & Ermold, W. (2001). PHC: A global ocean hydrography with a high-quality Arctic Ocean. *Journal of Climate*, *14*(9), 2079–2087. doi: Doi10.1175/1520-0442(2001)014⟨2079:Pagohw⟩2.0.Co;2

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., ... Roeckner, E. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172. doi: 10.1002/jame.20015

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ... others (2019). Dyamond: the dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1), 1–17.

Sundqvist, H., Berge, E., & KRISTJANSSON, J. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, *117*(8), 1641–1657.

Talley, L. D., Reid, J. L., & Robbins, P. E. (2003). Data-based meridional overturning streamfunctions for the global ocean. *Journal of Climate*, *16*(19), 3213–3226. doi: Doi10.1175/1520-0442(2003)016⟨3213:Dmosft⟩2.0.Co;2

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly weather review*, *117*(8), 1779–1800.

Tomita, H., Satoh, M., & Goto, K. (2002). An optimization of the icosahedral grid modified by spring dynamics. *Journal of Computational Physics*, *183*(1), 307–331.

Tomita, H., Tsugawa, M., Satoh, M., & Goto, K. (2001). Shallow water model on a modified icosahedral geodesic grid by using spring dynamics. *Journal of Computational Physics*, *174*(2), 579–613.

Wan, H., Giorgetta, M. A., Zängl, G., Restelli, M., Majewski, D., Bonaventura, L., ... others (2013). The icon-1.2 hydrostatic atmospheric dynamical core on triangular grids–part 1: Formulation and performance of the baseline version. *Geoscientific Model Development*, *6*(3), 735–763.

Wan, Z., Hook, S., & Hulley, G. (2015). *Mod11c1 modis/terra land surface temperature/emissivity daily l3 global 0.05deg cmg v006 data set.* distributed in netCDF format by the Integrated Climate Data Center (ICDC, icdc.cen.uni-hamburg.de) University of Hamburg, Hamburg, Germany. (last access date: January 6 2020)

Weller, H., Weller, H. G., & Fournier, A. (2009). Voronoi, delaunay, and block-structured mesh refinement for solution of the shallow-water equations on the sphere. *Monthly weather review*, *137*(12), 4208–4224.

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., ... Roeckner, E. (2019). *Mpi-m mpi-esm1.2-lr model output prepared for cmip6 cmip historical.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.6595` doi: 10.22033/ESGF/CMIP6.6595

Woodgate, R. A., Aagard, K., & Weingartner, T. J. (2006). Interannual changes in the Bering Strait fluxes of volume, heat, and freshwater between 1991 and 2004. *Geophys. Res. Lett.*, *33*, L15609. doi: 10.1029/2006GL026931

Woodgate, R. A., Weingartner, T., & Lindsa, R. (2012). Observed increases in Bering Strait oceanic fluxes from the Pacific to the Arctic from 2001 to 2011 and their impacts on the Arctic Ocean water column. *Geophys. Res. Lett.*, *39*, L24603. doi: 10.1029/2012GL054092

Zängl, G., Reinert, D., Ripodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 563–579. doi: 10.1002/qj.2378

Zhang, H.-M., Lawrimore, J. H., Huang, B., Menne, M. J., Yin, X., Sánchez-Lugo, A., ... Williams, C. N. (2019). Updated temperature data give a sharper view of climate trends. *Eos, Trans. Amer. Geophys. Union*, *100*. doi: 10.1029/2019EO128229