



Assessing resolution sensitivity in coupled climate simulations with AWI-CM3

Martina Zappaoni¹, Tido Semmler², Jan Streffing¹, Thomas Rackow³, Lettie A. Roach¹, and Thomas Jung^{1,4}

¹Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

²Met Éireann, 65–67 Glasnevin Hill, D09 Y921, Dublin, Ireland

³European Centre for Medium-Range Weather Forecasts (ECMWF), Bonn, Germany

⁴Department of Physics and Electrical Engineering, University of Bremen, 28359 Bremen, Germany

Correspondence: Martina Zappaoni (martina.zappaoni@awi.de)

Received: 15 October 2025 – Discussion started: 20 January 2026

Revised: 30 March 2026 – Accepted: 13 April 2026 – Published: 27 April 2026

Abstract. In this study, we evaluate the performance of the latest version of the Alfred Wegener Institute Climate Model, AWI-CM3, in two configurations at different resolutions, demonstrating that higher spatial resolution substantially enhances the model's ability to reproduce key climate variables and processes. The medium-resolution configuration consistently reduces climatological biases compared to both the low-resolution setup and the CMIP6 (Coupled Model Intercomparison Project, phase 6) multi-model mean, particularly in polar regions and areas characterized by strong mesoscale dynamics. Improvements are especially notable in the simulation of sea ice variability, ocean circulation, and ocean-atmosphere interactions. The medium-resolution simulation also exhibits greater interannual variability, which may reflect a more realistic representation of underlying processes, but whose implications will need to be fully assessed with multiple ensemble members. We conclude that long-term, eddy-permitting climate projections offer promising avenues for reducing structural uncertainties in future climate projections. As global modeling efforts move toward CMIP7 and beyond, our results highlight the importance of pursuing medium-resolution strategies in parallel with improved physical parameterizations and ensemble-based evaluation to more robustly capture the nonlinearities of the Earth system.

1 Introduction

The continuous development of climate models across successive phases of the Coupled Model Intercomparison Project (CMIP) progressively enhances our ability to simulate Earth's climate. However, studies have also highlighted persistent limitations in the current generation of models (Tokarska et al., 2020), raising concerns about the accuracy and reliability of their projections (Roach et al., 2020; Chen et al., 2021; Eyring et al., 2021; Slingo et al., 2022; Heuzé and Liu, 2024). These concerns are particularly relevant for the polar regions, where small-scale processes play a crucial role in shaping large-scale ocean and atmospheric dynamics (Heuzé et al., 2013; Marshall and Speer, 2012; Hewitt et al., 2022). One of the key factors contributing to these limitations is the relatively coarse spatial resolution of current climate models (Bock et al., 2020; Rackow et al., 2022; Li et al., 2024). In phase 6 of CMIP (CMIP6), global climate models (GCMs) typically use horizontal grid spacings of 100–200 km (roughly 1°) in both atmosphere and ocean components (Eyring et al., 2016). At these resolutions, critical mesoscale and submesoscale processes, such as eddies and turbulent mixing, are not fully resolved and must be parameterized, which indirectly affects other processes, including boundary currents, and introduces uncertainties into climate simulations (Hallberg, 2013; Griffies et al., 2015).

In the Arctic, CMIP6 models have proven capable of simulating atmospheric conditions with reasonable accuracy (Song et al., 2021; Zappaoni and Goessling, 2024), but exhibit biases in their representation of sea ice cover (Notz and SIMIP Community, 2020), sea ice-ocean interaction and

ocean circulation (Mulwijk et al., 2024; Athanase et al., 2025), as well as deep ocean convection (Heuzé et al., 2023). Regional evaluations show that CMIP6 models are broadly capable of capturing large-scale temperature and precipitation patterns in tropical to mid-latitude regions (Mmame and Ngongondo, 2024; Taylor et al., 2023), but biases persist in the representation of extreme events (Rabazanahary et al., 2024; Zhang et al., 2025a) and regions with complex orography (Pimonsree et al., 2023), emphasizing the importance of selecting appropriate models for regional climate studies. Additional systematic biases include the misrepresentation of the mid- to high-latitude cloud phase, potentially impacting the Earth's energy budget (Hellmuth et al., 2025). In the Southern Ocean, CMIP6 models struggle to accurately capture long-term sea ice trends (Roach et al., 2020; Casagrande et al., 2023) and short-term variability events such as polynyas (Mohrman et al., 2021). This is partially attributable to persistent biases in representing air-sea heat fluxes (Cesana et al., 2022, 2025). Similarly, the properties of the ocean water masses over the full depth remain poorly represented (Beadling et al., 2020). Taken together, these studies illustrate both the advances CMIP6 models have made over previous generations in simulating various high-latitude climate processes and the biases that remain, underscoring the need for model improvements, particularly for higher-resolution simulations capable of capturing small-scale processes and feedback mechanisms critical to polar climate dynamics.

Advancements in this direction have included the development of higher-resolution CMIP6 models, such as those in the HighResMIP protocol (Haarsma et al., 2016), where grid spacings are reduced to below 50 km in some cases. These simulations have demonstrated notable improvements in key aspects of climate representation, including ocean heat transport, atmospheric circulation, and sea ice distribution (Bock et al., 2020). More recently, coupled global simulations at kilometer-scale resolution (below 10 km) have become feasible, demonstrating striking improvements in the representation of mesoscale eddies and sea ice–ocean interactions (Hohenegger et al., 2023; Rackow et al., 2025; Moon et al., 2025; Doblas-Reyes et al., 2026).

However, while these higher-resolution models significantly reduce uncertainties associated with some parameterizations, their computational demands escalate rapidly with increasing resolution. Running models like this can require strategic partnerships, such as with the European High Performance Computing Joint Undertaking (EuroHPC) in the Destination Earth project (Doblas-Reyes et al., 2026). Global kilometer-scale ocean models can resolve mesoscale eddies across all latitudes and improve representations of the sea ice–ocean interface, including ice-shelf cavities. Their computational demand, however, generally limits them to decadal simulations, falling short of the timescales required for CMIP-class climate projections.

Here, we build on these developments by using the most recent version of the Alfred Wegener Institute (AWI) general circulation model, AWI-CM3 (Streffing et al., 2022), to perform fully coupled simulations spanning both the historical (1950–2014) and future (2015–2100) periods. We employ two distinct model configurations: a medium-resolution setup, MR (TCo319L137-DART) and a standard CMIP6-type low-resolution setup, LR (TCo95L91-CORE2). The nomenclature (e.g., TCo319L137-DART, TCo95L91-CORE2) reflects the horizontal and vertical resolution of the atmospheric component (spectral truncation TCo and vertical levels L) and the choice of ocean grid (DART or CORE2). The MR and LR experiments follow the CMIP6 HighResMIP protocol and include a 100-year spin-up to reduce model drift, ensuring comparability across resolutions.

The aim of this study is to assess how AWI-CM3's performance varies with spatial resolution, focusing on the impact of finer grid spacing on key climate variables, simulation accuracy, and the representation of small-scale processes. In this respect, our work is best understood as a companion piece to both Streffing et al. (2022) and Moon et al. (2025). Streffing et al. (2022) presented the first prototype of the model version used in all simulations analyzed here, primarily focusing on the configuration referred to in this study as LR. We extend this work by evaluating AWI-CM3 performance at a significantly higher resolution. In contrast, Moon et al. (2025) compared the MR configuration against a computationally demanding high-resolution (HR) setup, providing valuable insights into the benefits of kilometer-scale climate modeling, but over relatively short simulation periods and with an emphasis on selected processes. Here, we broaden the scope of analysis by also examining the general characteristics of the large-scale ocean and atmospheric circulation. In addition, our study compares MR directly with CMIP6-standard LR, assessing its added value for widely used intercomparisons. While Moon et al. (2025) emphasize the potential of HR simulations, we show that regionally enhanced simulations at MR provide long-term, computationally affordable simulations covering the full historical period and continuous projections through 2100, improving climate projections while remaining feasible for multi-century CMIP-class experiments.

2 Model setup

AWI-CM3 is a coupled global climate model consisting of the OpenIFS cycle 43r3 atmosphere model (ECMWF, 2017a, b, c) and the FESOM2 sea ice–ocean model (Danilov et al., 2017). Included in the atmospheric component are the WAM module for waves (Komen et al., 1996) and the H-TESEL module for hydrology (Balsamo et al., 2009). Coupled to it is the parallel XML Input/Output Server (XIOS) 2.5 (Yepes-Arbós et al., 2022). The ocean model incorporates the FESIM sea ice module (Danilov et al., 2015) which

employs a modified elastic–viscous–plastic (EVP) rheology (Koldunov et al., 2019b). AWI-CM3 also features a river routing scheme, which processes land surface water balance from the atmosphere and directs runoff to coastal discharge points of the ocean, based on river basin maps. More detailed descriptions of the model components and coupling can be found in Streffing et al. (2022).

The model is run using two different configurations: a lower and a higher resolution, called respectively LR and MR (medium-resolution) in this paper. The atmosphere model operates at the cubic octahedral TCo95 and TCo319 horizontal resolutions, which correspond to a range of approximately 102–120 and 32–38 km from the equator to the poles, with respectively 91 and 137 vertical levels up to a top pressure level of 0.01 hPa. The ocean model uses an unstructured mesh approach with resolution varying depending on the ocean dynamics and latitude (Sein et al., 2017, Fig. 1). Here, the horizontal resolution of the CORE2 mesh ranges from 20 km in the Arctic, 40 km across the Southern Ocean, and 100 km in subtropical regions (about 127 000 surface nodes, Koldunov et al., 2019a), whereas the DART mesh achieves finer resolution, from 4.5 km in the Arctic, 10 km across the Southern Ocean, and 25 km in subtropical regions (3.1 million surface nodes). In the vertical, the former has 47 vertical levels while the latter has 80 vertical levels. A more detailed description of the two meshes is available in Streffing et al. (2022). In both cases FESOM2 was configured with the linear free surface vertical coordinate system, which allows the ocean surface to move up and down in response to external forcing. In all the simulations reported here no partial cell layer at the bottom has been enabled.

The atmospheric model operates at 3600/900 s time steps, while the ocean model runs at 1800/240 s time steps for LR and MR, respectively. The coupled configuration employs the OASIS3-MCT_4.0 coupler (Craig et al., 2017) with a 2 h/1 h coupling time step. In Table 1 we report the computational performance for the two simulations.

The LR mesh does not resolve or permit ocean eddies (Fig. 1b, e); the effect of mesoscale eddies on tracers is included employing the Gent–McWilliams (GM) parameterization (Gent and McWilliams, 1990). In contrast, the MR mesh is eddy-permitting across most of the domain – except for the southernmost part of the Southern Ocean and regions with complex topography – and is locally eddy-resolving (Fig. 1c, f). As a result, the GM parameterization is disabled throughout the entire domain. Following standard modeling practice, the Redi scheme (Redi, 1982), which typically accompanies GM to represent isopycnal tracer diffusion, is also deactivated. This setup allows us to evaluate the historical performance and analyze future projections of a simulation that permits mesoscale dynamics across most of the ACC core and the inner Arctic Ocean. In both cases, ocean vertical mixing is simulated through the K-profile parameterization (KPP) mixing scheme (Large et al., 1994), but in the MR configuration an additional vertical mixing within the

Monin–Obukhov length scale (Timmermann and Beckmann, 2004) is enabled for the area south of 50° S to better represent vertical mixing under sea ice (Scholz et al., 2022).

2.1 Simulation protocol

As mentioned previously, both simulations were performed following the HighResMIP protocol, with an extension of the spin-up length to 100 years. A one-year coupled spin-up with a reduced ocean time step was performed beforehand exclusively for the MR configuration to mitigate initial shocks from the PHC3 ocean climatology initialization (PHC3.0, updated from Steele et al., 2001). From the final state of the spin-up, both a control simulation (ctl, 151 years) and a transient simulation (65 + 86 years) were initialized. During the spin-up and the control simulation, greenhouse gases and aerosol concentrations were fixed at 1950 levels. In contrast, the transient simulations followed the CMIP6 historical and SSP585 scenario forcings for greenhouse gases and aerosols (Meinshausen et al., 2017).

For both simulations, a single simulation has been performed and analyzed in this study. This is primarily due to the limitations with regard to available computing resources, which make the production of multiple ensemble members in the MR configuration challenging (Table 1). Further discussion of this aspect is provided in the Conclusions.

2.2 Model tuning

The two simulations were carried out using the same AWI-CM3 model version; however, as previously noted, the set of physical parameterizations applied varies with spatial resolution. The goal of this study is to evaluate the impact of resolution on model performance. To achieve this we have chosen to compare two simulations that are each optimized for their respective grid resolution. Parameter choices in both setups result from dedicated tuning processes designed to maximize performance at each resolution, which closely follows the approach of the HighResMip2 protocol for CMIP7 (Roberts et al., 2025). Differences in parameter settings were anyway kept to the minimum, to ensure as clear a comparison as possible, with only a small number of technical parameters adjusted during tuning. Among the tuned parameters, albedo usually plays a particularly influential role due to its strong effect on the model's radiative balance and surface energy fluxes. In our model, distinct four constant albedo values are prescribed for sea ice, snow, and melting ice or snow. A common set of values was identified during the tuning process and applied consistently across both configurations.

It is worth noting that the simulation presented in this study is based on a slightly updated version of AWI-CM3 compared to the setup described in Moon et al. (2025), though the two remain highly comparable. The most significant difference concerns the remapping method used for ice-related heat fluxes exchanged between the atmosphere and

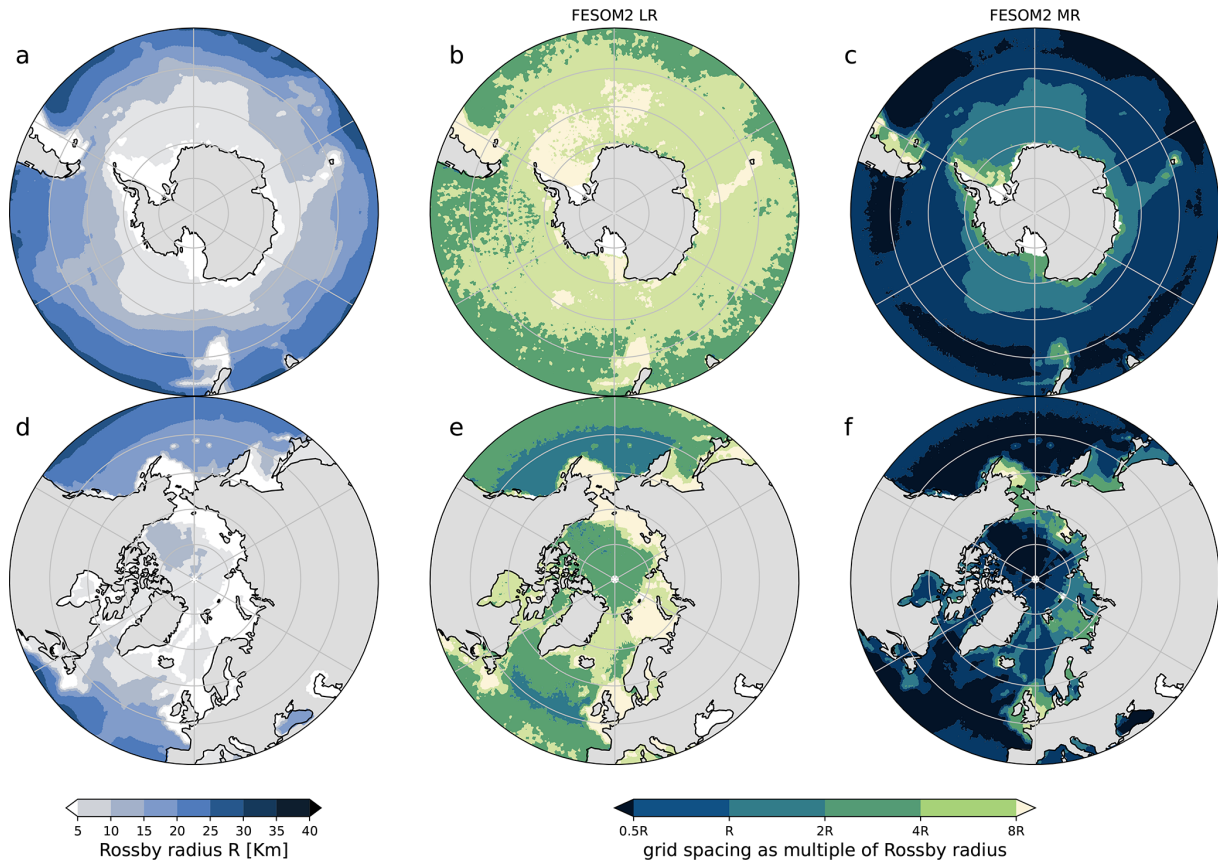


Figure 1. Local Rossby radius of deformation (**a, d**) estimated as $R = (\pi f)^{-1} \int_{-H}^0 N(z) dz$ where f is the Coriolis parameter, H is the ocean floor, and N is the Brunt-Väisälä frequency (as in Sein et al., 2016). Ocean horizontal grid spacing as a multiple of the local Rossby radius for the (**b, e**) low and (**c, f**) high resolution meshes (CORE2 and DART respectively). Blue (dark blue) colour marks the eddy-permitting (eddy-resolving) areas.

Table 1. AWI-CM3 computational performance in the two different configurations. Atmospheric grids and ocean meshes are listed together with their respective horizontal resolution ranges (in parentheses). The total number of cores used, the number of simulated years per day (SYPD) and the core hours per simulated year (CHSY) are reported for both configurations.

	Atmospheric grid	Ocean mesh	IO scheme	Cores	SYPD	CHSY
LR	TCo95L91 (102–120 km)	CORE2L47 (20–100 km)	XIOS parallel	1664	73.80	541
MR	TCo319L137 (32–38 km)	DARTL80 (4.5–25 km)	XIOS parallel	8576	3.89	52 911

ocean via the model coupler. In the earlier setup, the Gaussian remapping method introduced a spurious heat flux over ice-covered regions of the Northern Hemisphere leading to a reduced sea ice extent and premature disappearance of sea ice (Fig. A1). In this model version that has been replaced by a bi-cubic remapping method that improves the Arctic sea ice representation. The Southern Hemisphere was not affected by the previous issue, and we do not observe a substantial change in the simulation of Antarctic sea ice.

3 Results

3.1 Model drift and equilibration

Before evaluating the climatological performance of the simulations, we assess the degree of equilibration of the coupled system under constant forcing. While AWI-CM3-LR exhibits a clearly positive top-of-atmosphere (TOA) radiative imbalance since the beginning of the simulation, AWI-CM3-MR starts with near-zero values; both simulations, however, display a declining trend during the spin-up phase (Fig. 2a, b). After spin-up, the imbalance stabilizes at approximately $+0.2 \text{ W m}^{-2}$ in the low-resolution configura-

tion, reasonably on the lower of CMIP6 models average and reference estimates for the present-day climate (Johnson et al., 2016; Wild, 2020), and at a nearly symmetric negative value ($\sim -0.24 \text{ W m}^{-2}$) in the medium-resolution case. The surface energy imbalance is smaller in the LR simulation and more pronounced in the MR configuration, but in both cases it closely follows the evolution of the TOA flux. The relatively stable TOA–surface difference indicates that the atmospheric column remains internally consistent, while suggesting that, particularly in the MR simulation, the residual energy imbalance is primarily associated with ongoing ocean adjustment.

In both configurations, the global-mean ocean temperature bias exhibits a relatively stable vertical structure (Fig. 2c, d). However, differences emerge in both the temporal evolution and the magnitude of these biases. In the LR simulation, a weak cold bias around 100 m is present during the initial adjustment phase and diminishes over time as warm biases at intermediate depths become more pronounced. This is accompanied by a persistent subsurface warming throughout the control period, indicating a sustained drift in the ocean interior despite the near-stationary behavior of the global energy budget, broadly consistent with Streffing et al. (2022). In contrast, the HR simulation shows a more stable depth-dependent pattern and less pronounced ocean temperature drift, with weaker subsurface warming and a clear upper-ocean cooling, consistent with a net upward energy flux at both the surface and the TOA.

While the low-resolution simulation can be considered reasonably equilibrated by 1950, the medium-resolution configuration still exhibits signs of residual imbalance at the time the transient simulations are branched off. This is reflected in the larger variability of the radiative balance and is consistent with the presence of an eddy-permitting ocean. Explicitly resolving mesoscale processes substantially increases the timescale required for the deep ocean to adjust (Danabasoglu et al., 1996). As a result, the approach to equilibrium is slower, and residual imbalances can persist over longer periods compared to lower-resolution configurations relying on parameterized processes (Danek et al., 2019).

Contrasts between low- and high-resolution configurations have been reported in HighResMIP and other multi-model studies, where increasing resolution leads to a reorganization, or even a reversal in sign, rather than a systematic reduction of model errors (Bock et al., 2020; Moreno-Chamarro et al., 2022). However, most relevant for the present comparison is that both simulations exhibit an overall limited and well-constrained drift from 1950 onwards. This is evident in the weak trends of global-mean sea surface temperature and near-surface air temperature, as well as in the relatively stable sea ice extent across seasons and regions (Fig. 4).

3.2 Historical climatology

To obtain an objective overview of the overall performance of the model, we use a multivariate comparison against observational climatologies. This approach provides a broad assessment of how well the model captures key climate features with respect to the CMIP6 multimodel ensemble. Eighteen key climate variables were selected for this evaluation. For each variable, we calculate the absolute climatological error in the AWI-CM3 simulations, considering all four seasons and seven distinct regions. The absolute error is first computed at each grid cell of an interpolated grid and is then averaged over each of the seven regions and over each season. Finally, the resulting model error is compared to the average absolute error of 30 CMIP6 models and expressed as a fraction of this CMIP6 mean error. Values below 1 indicate better performances, whereas values above 1 suggest larger errors (Fig. 3a). Only the last 25 years of the historical simulation (from December 1989 to November 2014, in line with the seasonal analysis) have been considered here to maximize comparability with the observational datasets. It is worth noting that not all observational datasets cover the entire period; a detailed overview of the coverage period for each variable is provided in Streffing et al. (2025). To directly compare the two simulations, the index for AWI-CM3-MR was computed again using AWI-CM3-LR as the simulation reference for the evaluation (Fig. 3b). In this way the MR simulations show better performances than the LR simulations for values below 1 and larger errors for values above 1.

The overall score, computed as the average across all variables, regions, and seasons, indicates a general improvement over the CMIP6 standard (Figs. 3a and A2) which, in the MR configuration (AWI-CM3-MR), corresponds to an average error reduction of 27.9%. Examining the variable-specific errors, the simulation demonstrates a widespread reduction in the atmospheric absolute error relative to the CMIP6 models considered here, along with notable improvements in ocean dynamics. Large errors in sea surface temperature (tos) are primarily observed in the mid- and tropical latitudes and are associated with a persistent warm bias in the model (discussed in detail in Sect. 2.2), which is significantly reduced at higher latitudes, particularly in the Southern Hemisphere, when resolution is increased. It was previously noted that the model tends to display a substantial misrepresentation of mixed-layer depth (mldst) in the Antarctic (Streffing et al., 2022), however in this case it is only in the MR configuration showing this error. Despite the larger error magnitude with respect to the LR configuration, the MR setup produces more realistic sea ice concentration (siconc) values in the region in austral summer and autumn, but shows less agreement with observations in winter and spring. In this case, the comparison among simulations is complicated by the notably higher variability of the MR simulation during the analysis period. In the Northern Hemisphere, sea ice concentration performs well across all seasons, except for an overestimated

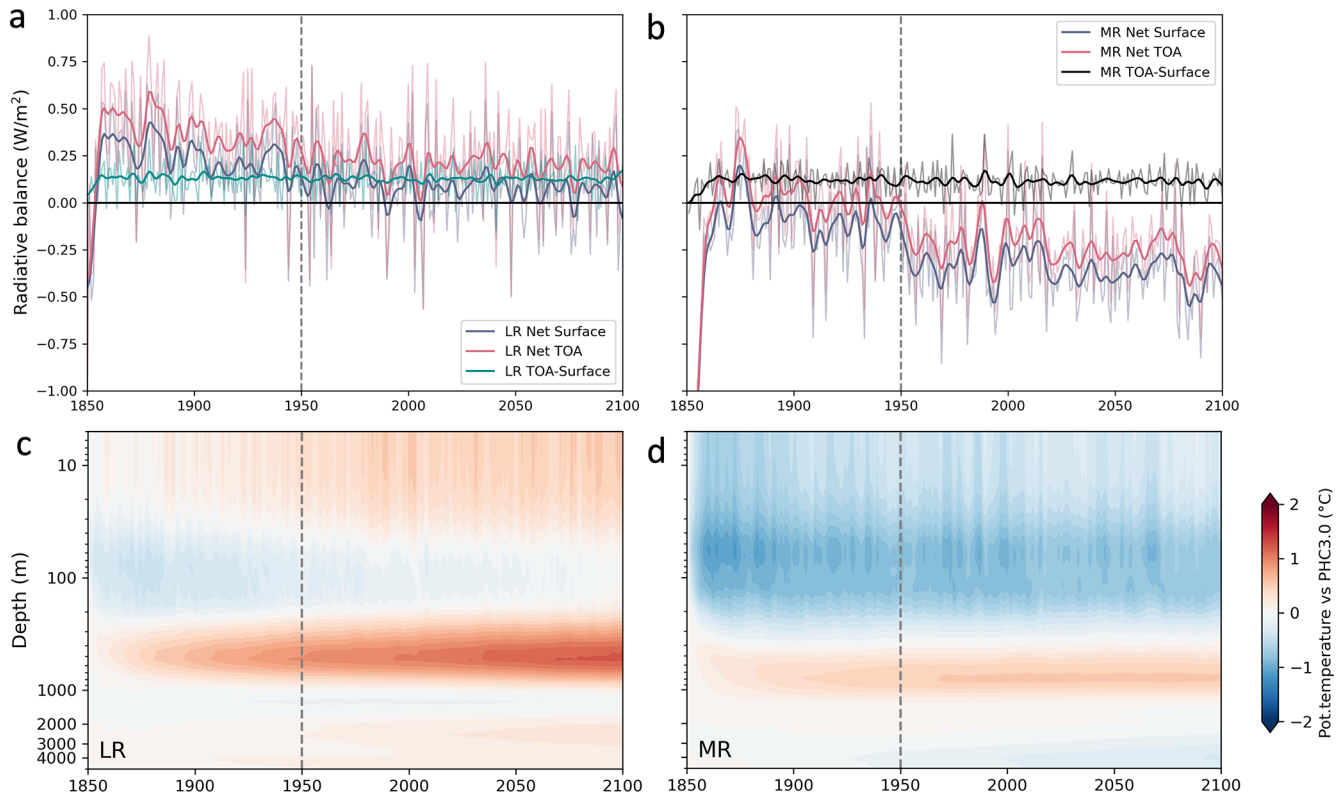


Figure 2. (a, b) Net radiative imbalance at the top of the atmosphere (TOA) and at the surface, together with their difference, representing the atmospheric column energy balance, during the spin-up and control simulations. Positive (negative) values indicate downward (upward) fluxes. (c, d) Hovmöller diagrams of the evolution of global-mean ocean potential temperature bias during the spin-up and control simulations, relative to the PHC3.0 climatology. The vertical dashed line marks the branch-off point of the transient simulations. Results are shown for the low-resolution (LR, left panels) and medium-resolution (MR, right panels) configurations.

summer sea ice extent at mid-latitudes (discussed in details in Sect. 2.3), an error that is most pronounced in the MR simulation. The 2 m temperature is generally well-simulated, here the increased resolution leads to more substantial improvements in the Antarctic compared to the Arctic. Higher spatial resolution also has a broadly positive impact on surface winds (u_{as} and v_{as} , representing the zonal and meridional components respectively) at the global scale. Similar improvements are observed for the 300 hPa zonal wind (u_a), the 500 hPa geopotential height (z_g), and the top-of-atmosphere outgoing longwave radiation ($rlut$). Precipitation (pr) is improved primarily in the tropics, while enhancements in cloud coverage (clt) are most evident at high latitudes. Ocean temperature and salinity (θ_{tao} and so) are, on average, well simulated; however, the impact of increased resolution is not uniform across regions or ocean depths. The higher resolution in MR generally helps reduce the ocean warm bias, except in southern mid-latitudes at 1000 m depth, where the MR is warmer (Fig. A3a–d). On the other hand the salinity bias, which is still below average in the LR simulation (Fig. A2), is amplified at the surface, in the tropical upper ocean and in the Southern Ocean at 1000 m depth (Fig. A3e–h).

The overall evaluation scores show that the LR simulation achieves a $\sim 13\%$ error reduction relative to the CMIP6 multi-model mean, while the MR simulation improves this further to a $\sim 28\%$ reduction. Therefore, approximately 55% of the total error reduction compared to the CMIP6 standard can be attributed to the increase in spatial resolution. While this overall score provides a useful summary of model performance, it involves an inherent degree of arbitrariness, as it depends on the selection of variables and regions. It is therefore interpreted alongside the individual diagnostics presented below, rather than as a standalone measure of model skill. In the following, a more detailed analysis is conducted to assess specific aspects of the two simulations.

3.3 Surface temperature

A visible difference between the two model configurations lies in their mean state temperature bias. Even in the control simulation, a persistent offset is evident: the LR configuration is, on global average, approximately 1.5°C warmer than the MR configuration for both sea surface and near-surface air temperatures (Fig. 4a). A similar bias is also evident in the transient low-resolution simulation when compared

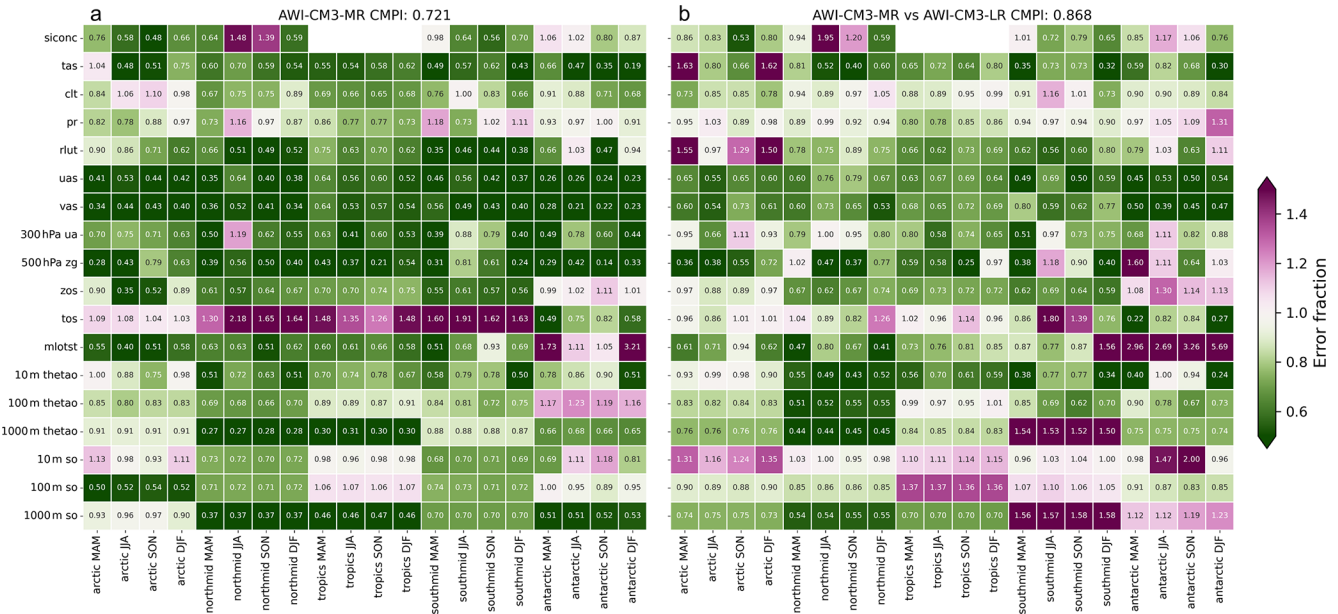


Figure 3. (a) Performance indices showing the absolute error of the AWI-CM3-MR historic climatology (1989–2014) expressed as a fraction of the CMIP6 multi-model mean absolute error (30 models; Streffing et al., 2025; Reichler and Kim, 2008). Values below (above) 1 indicate smaller (larger) biases compared to the CMIP6 average. Panel (a) compares AWI-CM3-MR against CMIP6, while Fig. A2 presents the same comparison for AWI-CM3-LR. (b) Performance indices comparing AWI-CM3-MR with AWI-CM3-LR, using the LR as reference. Here, values below (above) 1 indicate a reduction (increase) of error in MR relative to LR. The Arctic and Antarctic regions are defined as latitudes north of 60° N and south of 60° S, respectively; the mid-latitudes span 30–60° in both hemispheres; and the tropics cover 30° S–30° N. The list of all variables with full names is reported in Table A1.

to observations, becoming particularly pronounced from the last decade of the 20th century onward. In contrast, the MR setup more closely reproduces the observed magnitude of the global mean surface temperature. During most of the satellite era (1990 onward), the MR configuration better represents the observed near-surface air temperature trend (MR: 0.25 °C per decade; LR: 0.19 °C per decade; ERA5: 0.24 °C per decade).

For sea surface temperature, both simulations exhibit a warming trend stronger than observed, with the MR configuration showing the largest deviation (MR: 0.18 °C per decade; LR: 0.14 °C per decade; HadISST: 0.11 °C per decade). It is also worth noting that the low-resolution simulation exhibits a stronger residual model drift in its control experiment, which is substantially mitigated by increasing the spatial resolution, consistent with the suggestion of Griffies et al. (2015), who highlight the role of numerical methods and physical parameterizations in contributing to model drift. Over the 151-year simulations, the total surface temperature drift in the LR configuration amounts to 0.21 °C in the ocean and 0.24 °C in the atmosphere, compared to only 0.06 and 0.10 °C in the MR configuration, respectively.

The spatial distribution of the biases highlights how the surface warm bias in the lower-resolution setup is fairly homogeneous at the global scale, affecting both atmosphere and ocean (Fig. 5a, i). Nonetheless, specific hot spots emerge –

such as the El Niño/La Niña upwelling regions, the Southern Ocean, and the so-called northwest corner of the North Atlantic Ocean, identified as common biases in GCMs (Sein et al., 2017; Bock et al., 2020; Streffing et al., 2022) and references therein) – as well as the Kuroshio region. Consistent with comparisons between CMIP6 and HighResMIP (Bock et al., 2020), the magnitude of these biases is significantly reduced when horizontal resolution is increased (Figs. 5e, m and A4). The negative surface temperature bias near the equator, previously identified in the AWI-CM3 prototype (Streffing et al., 2022), is relatively weak in this LR simulation and is mostly confined to land areas. Conversely, much of the low- to mid-latitude region exhibits a cold bias in the higher-resolution configuration. The 2 m temperature bias over the central Arctic can to a large degree be explained by a warm bias present in the ERA5 reanalysis (Tian et al., 2024). At higher resolution, this negative bias exceeds 4 °C and peaks in the Canadian sector of the Arctic Ocean, while it is less pronounced at lower resolution. This pattern also resembles results obtained from comparisons of CMIP6 and HighResMIP multi-model means (Bock et al., 2020). However, as mentioned above, this difference must be contextualized, as recent studies indicate that ERA5 itself exhibits a warm bias exceeding 3 °C in the same region when compared with satellite observations (Tian et al., 2024).

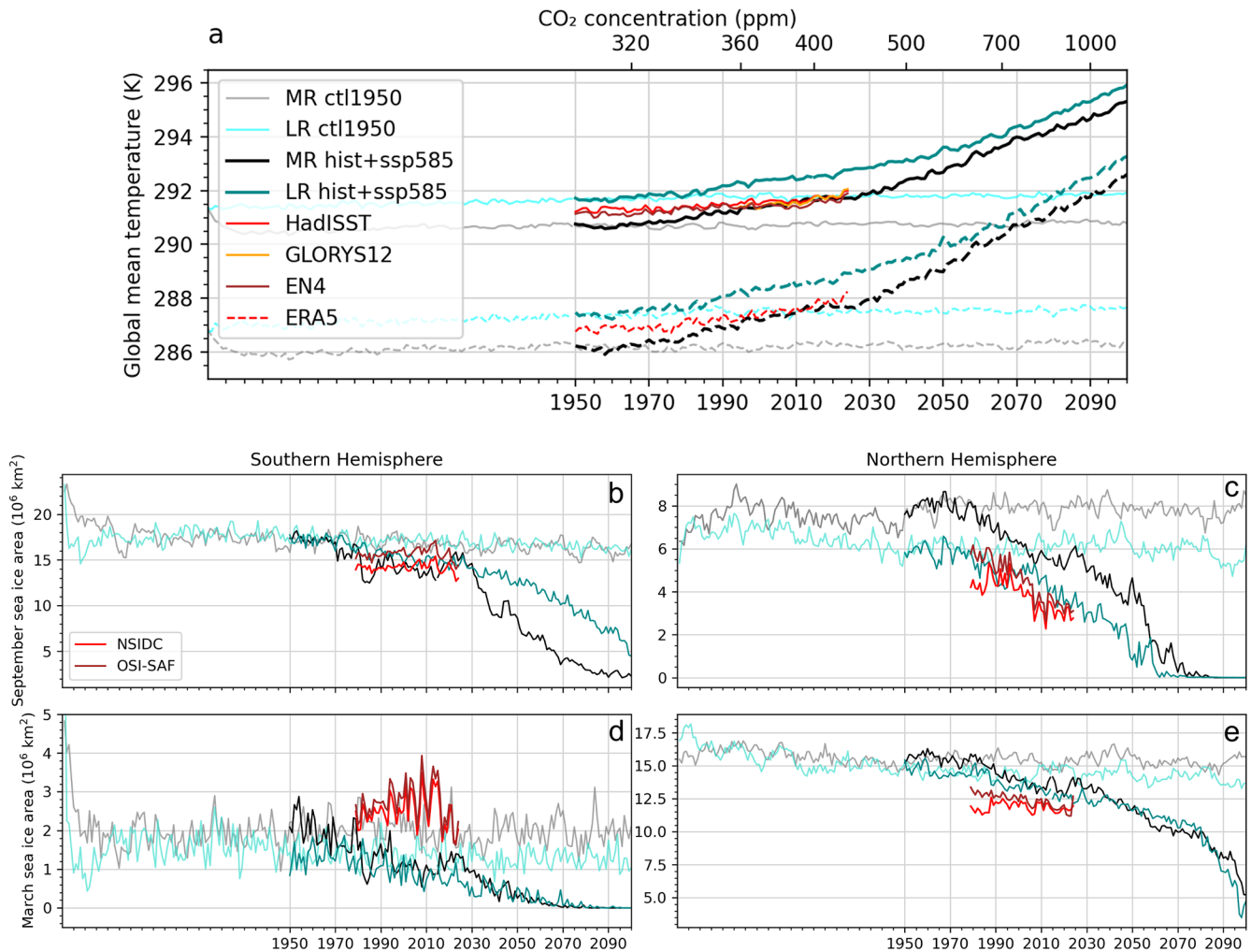


Figure 4. Time series of (a) global mean sea surface (full line) and 2 m air (dashed line) temperatures and (b–e) sea ice extent for the two hemispheres. Temperature is reported as annual mean, while for sea ice September and March means are represented separately. Light/dark green represent the low resolution experiment (LR) and gray/black lines represent the medium resolution experiment (MR). Light green and gray show the control runs while dark green and black show the hist+ssp experiments. Due to observational uncertainty, especially when looking at polar regions, multiple datasets have been included for observed sea surface temperature, 2 m air temperature, and sea ice area: HadISST1 (Rayner et al., 2003), GLORYS12 (Lellouche et al., 2021), EN4 (Good et al., 2013), ERA5 (Bell et al., 2021), NSIDC (DiGirolamo et al., 2017) and OSI-SAF (Lavergne and Down, 2023).

The analysis of the last 35 years of the historical simulations indicates that the MR experiment exhibits a substantially reduced mean temperature bias, decreasing from 0.85 to -0.24 °C for global 2 m air temperature, and from 0.88 to -0.05 °C for sea surface temperature. Moreover, the observed global warming trend during this period is well reproduced in the MR simulation. Examining regional trends in the extratropics (Fig. 5b–d, f–h, j–l, n–p), the benefits of increased resolution are still generally evident, though in some regions less pronounced. In the Southern Hemisphere, the LR simulation fails to capture the observed cooling trend, instead displaying widespread warming in both the atmosphere and ocean (Fig. 5c, k). The MR simulation, by contrast, introduces more spatial detail: coastal regions tend to show

weaker positive or neutral trends, and localized cooling, especially in the Pacific sector, is better represented (Fig. 5d, l). In the Northern Hemisphere, the MR simulation produces an overly strong surface air warming trend—especially pronounced at mid-latitudes—when compared to reanalysis data, likely related to the more pronounced cold bias (Fig. 5h). Meanwhile, the LR experiment shows only modest warming across much of the Arctic Ocean, with stronger trends confined to the Greenland and Barents Seas (Fig. 5g). For sea surface temperatures, both simulations reveal broadly similar warming patterns in the Northern Hemisphere extratropics; the erroneous negative trend in the Labrador Sea is here more clearly visible (Fig. 5o, p), especially in the MR simulations. Overall, while the MR simulation significantly reduces

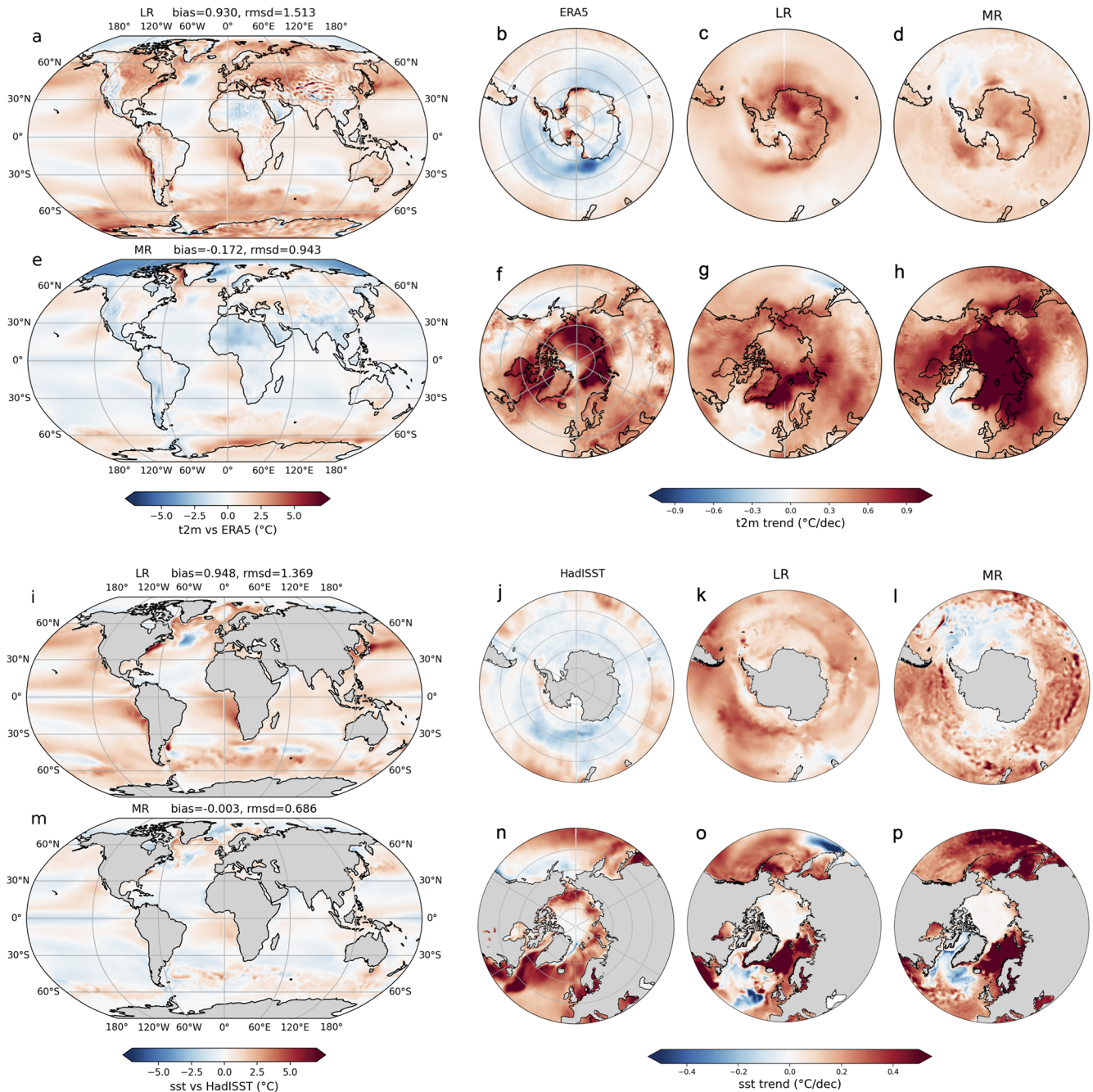


Figure 5. Annual mean (a, e) 2 m air temperature and (i, m) sea surface temperature biases for the period 1979–2014 relative to ERA5 and HadISST, respectively. Bias and rmsd represent the area-weighted mean and root mean square difference between model and observations. Annual mean 2 m air temperature (b–d, f–h) and sea surface temperature (j–l, n–p) trends over the same period, separately for the Northern and Southern extratropics (regions poleward of 40° N and 40° S), for both observations and model simulations. Results are shown for both the low-resolution (LR) and medium-resolution (MR) experiments.

global mean temperature biases and better captures regional details, especially in the Southern Hemisphere, its improvements in replicating observed spatial trend patterns remain limited and regionally dependent for this individual realization. It is worth noting that regional trend patterns can be strongly affected by internal variability and part of the mis-

match with observations may reflect this variability rather than systematic model biases.

3.4 Sea ice and mixed layer depth

The medium-resolution configuration exhibits a stronger increase in air temperature per unit of CO₂ concentration already during the historical period, with this difference becoming even more pronounced starting from the late 2020s (Fig. 4a). A similar, though slightly less pronounced, signal is also evident in the sea surface temperature. It is noteworthy that this year also marks the onset of a pronounced regime shift in the maximum sea ice area over the Southern Ocean (Fig. 4b): within approximately 15 years, Antarctic sea ice loses nearly 7 million km² of its area. This rapid decline is analyzed in detail by Kim et al. (2026). Here, we simply note that such a sharp decline is absent in the low-resolution simulation, which instead shows a more gradual and stable decrease typical of many CMIP6 models. Enhanced sea ice variability in the Southern Ocean is also apparent prior to this event in the medium-resolution simulation, with several fluctuations linked to more or less severe Weddell Sea polynya episodes. For this reason, despite being an average over a long period of 25 years, the error displayed when computing the performance index (Fig. 3a) is influenced by this strong variability.

Several studies have shown that CMIP6 climate models tend to simulate a stably stratified Southern Ocean (Mohrman et al., 2021). Consistent with this, the low-resolution AWI-CM3 simulation exhibits a shallow winter mixed layer depth (MLD) with very limited variability across most of the region south of 60° S (Fig. 6a, c). Notably, the Weddell Sea lacks any significant deep MLD features, showing some deepening only in a narrow coastal region. This behavior aligns with the limited polynya features in this model configuration. In contrast, the medium-resolution AWI-CM3 simulation shows pronounced MLD variations (Fig. 6b, d). Along the Antarctic coast, the MLD exceeds 200 m, consistent with active coastal polynya formation. In the Weddell Sea, the MLD reaches depths greater than 300 m, making it the region with the strongest variability in the Southern Ocean. This deeper and more variable MLD contributes to a more plausible representation of polynya occurrence, while at the same time leading to an overestimation of event intensity. It is important to note, however, that the observational record of polynya events is very limited, which adds uncertainty to such evaluations. The overly deep mixed layer is collocated with a localized and pronounced warm bias in 2 m air temperature (Fig. 5e), likely linked to enhanced oceanic heat release in this region. In contrast, the low-resolution configuration shows no such localized warm bias, instead displaying a relatively uniform positive bias across the Southern Ocean.

This discrepancy explains why the medium-resolution simulation exhibits greater variability in the inner Weddell Sea sea ice maximum compared to observations (Fig. 7a, e). Modeled September sea ice concentration trends in the Southern Hemisphere mirror the corresponding temperature trends (Figs. 5 and 7c, e): while the low-resolution simula-

tion shows a largely uniform decline, the medium-resolution experiment captures some localized increases, such as in the Ross Sea and the Weddell Sea. However, both simulations display a marked retreat in sea ice concentration across East Antarctica, an inconsistency with satellite observations. In the Northern Hemisphere, the observed boreal summer sea ice loss in the central Arctic is largely absent in the low-resolution run but is reproduced, albeit with regional biases, at higher resolution, including excessive retreat in the Greenland and Barents Seas and an absent response in the Russian-Canadian sector (Fig. 7g, i, k). In March, sea ice extent is overestimated in the medium-resolution simulation (Fig. A51) and shows indeed a more spatially distributed decline, stronger particularly at mid-latitudes (Fig. 71), consistent with stronger positive temperature trends. In contrast, the low-resolution simulation exhibits spatial variability that more closely aligns with satellite data (Fig. 7i). During austral summer, neither simulation successfully reproduces the observed sea ice trends of the Southern Hemisphere (Figs. 7b, d, f and A5b, d, f). Overall, the main improvement associated with reduced grid spacing is a better representation of average sea ice trends (Fig. 4) and an enhanced ability to capture regional variability in the Southern Hemisphere. However, both resolutions still fall short in reproducing key observed features, such as the stability of East Antarctic sea ice and the full extent of Arctic summer sea ice loss.

3.5 Ocean circulation

The AWI-CM3 simulations reproduce the expected structure of the AMOC streamfunction in its mean state (Fig. 8a, b). However, two main differences emerge when comparing the streamfunctions from the low- and medium-resolution simulations. In the upper cell, within the 500–1500 m depth range, the low-resolution simulation shows a peak in northward transport primarily between 30 and 40° N, whereas the medium-resolution simulation displays a more uniform maximum across the latitudes. The comparison with observations at 26° N reveals that the medium-resolution simulation aligns more closely with the RAPID profile in the upper ocean but tends to overestimate the southward branch of the AMOC, suggesting a too-strong return flow of North Atlantic Deep Water (NADW) along the Deep Western Boundary Current (DWBC) (Fig. 8c, d). Conversely, the low-resolution simulation, while better capturing the magnitude of the NADW transport, locates its flow too shallow compared to observations (Fig. 8c, d), consistent with findings by Gutjahr et al. (2019), and with a too pronounced variability. Additionally, the medium-resolution simulation shows a clearer signal of northward transport of Antarctic Bottom Water (AABW) at abyssal depths (Fig. 8a, b). Similar resolution-dependent differences have been reported in earlier studies (e.g., Li et al., 2022).

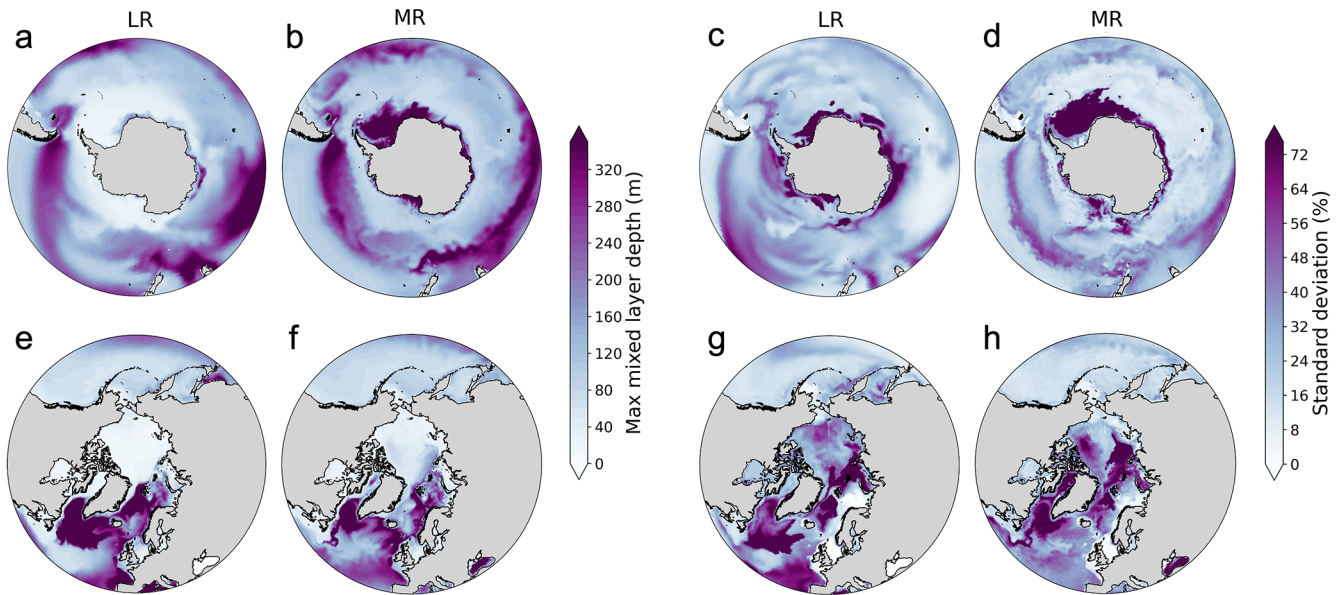


Figure 6. 1979–2014 maximum winter mixed layer depth (a, b, e, f) and standard deviation (c, d, g, h) for the low resolution (LR, a, e, c, g) and medium resolution (MR, b, f, d, h) simulations. Respectively austral/boreal (JJA/DJF) winter season have been considered for the Southern/Northern hemisphere. MLD is defined as the depth at which the potential density differs by 0.03 kg m^{-3} from the surface density (Monterey and Levitus, 1997).

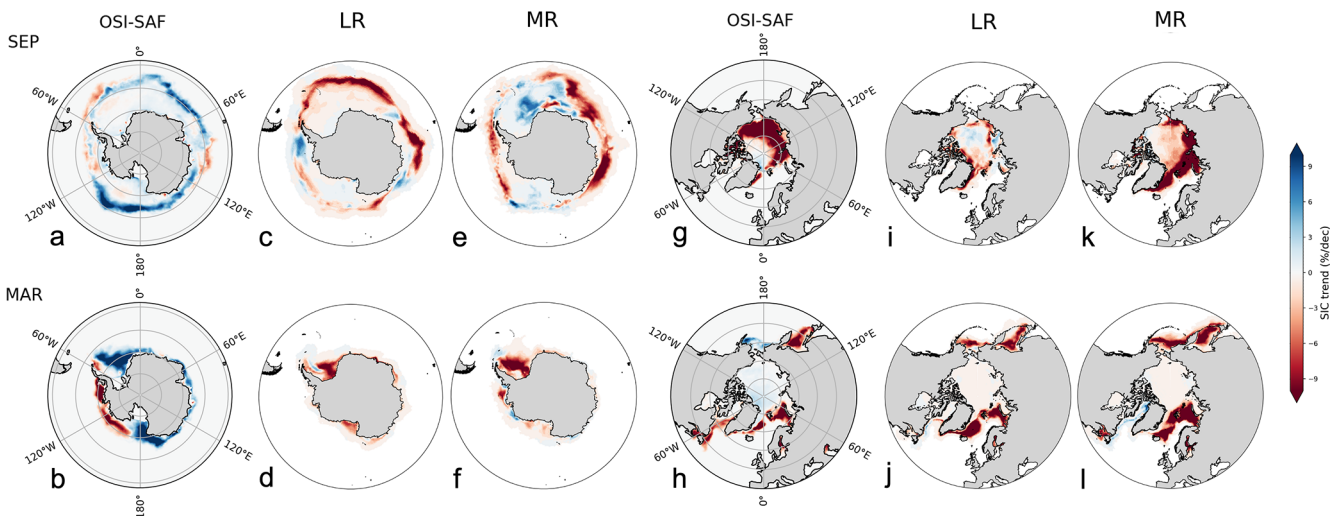


Figure 7. Trends in March and September sea ice concentration from 1979 to 2014 are shown for both hemispheres, based on OSI-SAF observations (a, b, g, h) and simulations from the low-resolution (LR, c, d, i, j) and medium-resolution (MR, e, f, k, l) model configurations. Although the two observational datasets display some differences in the time series shown in Fig. 4b–e, their spatial patterns of sea ice trends are largely consistent. The decision to present only OSI-SAF data here is motivated by its treatment of the central Arctic, where the typical data gap has been filled through interpolation.

Starting around 1980, both simulations show a decline in the AMOC at 26.5°N under increasing atmospheric forcings (trend absent in the control simulations), but this weakening is confined to the northward flow component of the upper cell. In contrast, the southward branch exhibits no clear trend in the low resolution simulation, indicating a deep ocean circulation that is less sensitive to surface-driven

changes, at least initially during the transient response phase. On the other hand, higher resolution shows more coherent changes between northward and southward transport, showing a decline in DWBC especially pronounced in the last two decades of the 21st century. As shown in Putrasahan et al. (2019), improving atmospheric resolution can lead to significant AMOC weakening via salinity-driven suppression of

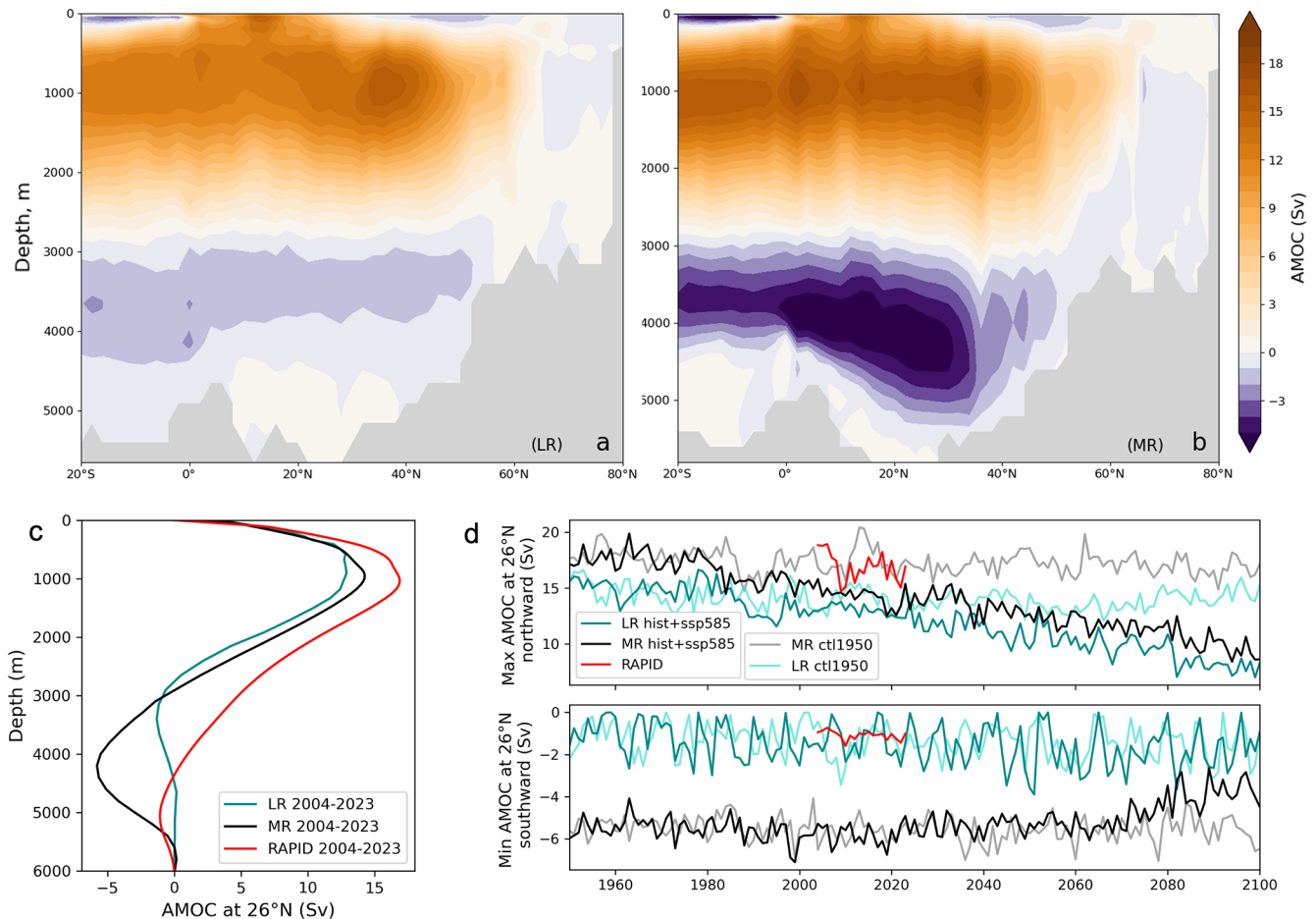


Figure 8. Atlantic Meridional Overturning Circulation (AMOC) overturning streamfunction averaged over the period 1979–2014 for (a) the low-resolution simulation (LR) and (b) the medium-resolution simulation (MR). (c) Comparison of AMOC overturning streamfunctions from AWI-CM3 simulations and RAPID-MOC array (Moat et al., 2024) observations at 26.5° N, averaged over the period 2004–2023. (d) Time evolution of AMOC maximum (northward transport) and minimum (southward transport) at 26.5° N from AWI-CM3 simulations and RAPID observations. (Sv denotes Sverdrups; $1 \text{ Sv} = 106 \text{ m}^3 \text{ s}^{-1}$).

deep convection in the North Atlantic deriving from altered wind and ocean circulation.

Comparing the simulated volume transport across major ocean channels with observational estimates, provides initial insight into the effects of both increased horizontal resolution and a finer vertical discretization on ocean circulation and vertical processes (Table 2). In the Arctic, the medium-resolution simulation generally produces stronger volume transports across key straits compared to the low-resolution setup, with a tendency toward a positive bias relative to observations. Through the Davis and Fram Straits, the medium-resolution simulation falls within the observational uncertainty range, while the low-resolution one tends to underestimate the outflow. In contrast, the medium-resolution simulation overestimates poleward transports through the Barents and Bering Seas, leading to an excessive inflow of warm Atlantic and Pacific waters. This increased inflow likely contributes to a positive near-surface salinity bias in the Arctic

Ocean, which is more pronounced in the medium-resolution case (Fig. 9a–d). This bias is consistent with known shortcomings of AOGCMs, often attributed to an insufficient freshwater supply from Siberian rivers (Gutjahr et al., 2019). At the same time, the enhanced northward transport in the medium-resolution simulation slightly reduces the warm bias (Fig. 5a, e) and significantly decreases the salinity bias in the North Atlantic near-surface (Fig. 9a, b).

Improvements are also evident in the representation of the Gulf Stream separation, which reduces the localized fresh bias in the North Atlantic surface, and in the freshening of the Mediterranean Sea, which mitigates the excessive outflow of overly salty and warm Mediterranean waters around 40° N (Fig. 9c–f), present in the LR. These Mediterranean waters, in the low-resolution simulation, spread northward along the European continental shelf, contributing to a saltier and warmer NADW, which is then transported southward. The medium-resolution simulation shows improvements in

Table 2. Mean ocean volume transport through selected straits and channels as simulated by AWI-CM3 in both configurations, compared with observational estimates (positive means northward or eastward). Model values represent averages over the period 1990–2014 of the historical simulation. References correspond to observational studies cited for each strait. (Sv denotes Sverdrups; $1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$).

Transport (Sv)	LR	MR	Observations
Davis Strait	−0.63	−1.1	-1.6 ± 0.5 (Curry et al., 2014)
Fram Strait	−1.74	−2.65	-2.2 ± 2.1 (Beszczynska-Möller et al., 2015)
Barents Sea Opening	2.33	4.27	2.1 ± 0.4 (De Boer et al., 2018)
Bering Strait	1.4	1.6	0.8 ± 0.2 (Woodgate et al., 2005) 1.00 ± 0.05 (Woodgate, 2018)
Straits of Gibraltar	−0.24	0.044	0.038 ± 0.007 (Soto-Navarro et al., 2010)
Makassar Strait (ITF)	−7.39	−8.71	−11.6 (Gordon et al., 2010)
Mozambique Channel	−38.68	−26.43	-16 ± 8.9 (Ridderinkhof et al., 2010)
Drake Passage	215.64	199.05	136.7 ± 6.9 (Meredith et al., 2011) 173.3 ± 10.7 (Donohue et al., 2016)

ocean circulation and mixing processes, leading to a more realistic representation of deep water formation and large-scale circulation in the North Atlantic Ocean, consistent with the findings of Rackow et al. (2019).

Consistent with Griffies et al. (2015), resolving mesoscale eddies in the medium-resolution simulation reduces the positive deep-ocean temperature bias overall, indicating more efficient vertical heat transport from the interior—with the exception of 1000 m depth where MR shows a slightly warmer bias than LR simulation in the mid-latitudes of the Southern Hemisphere (Fig. A3c, d). However, the surface ocean change from LR to MR is different from the classical eddy-heating effect in Griffies et al. (2015), where SST tends to warm when explicitly resolving mesoscale eddies. This pattern suggests that, while mesoscale processes improve deep-ocean realism, their effect on subsurface and surface temperatures is spatially heterogeneous, likely reflecting regional feedbacks such as stratification, sea ice dynamics, and coupled ocean-atmosphere interactions.

At higher southern latitudes, both model configurations tend to overestimate volume transports through the selected channels (Table 2). However, the medium-resolution simulation generally falls closer to the observational uncertainty ranges. A clear example is the transport through the Drake Passage, where both simulations produce a stronger Antarctic Circumpolar Current (ACC) than older observational estimates (e.g., Meredith et al., 2011). More recent full-depth measurements that include both baroclinic and barotropic components of the flow, however, indicate higher transport values (Donohue et al., 2016). In this context, the AWI-CM3 medium-resolution configuration approaches the upper bound of these updated observational estimates.

A well-known bias in low-resolution models is the incorrect representation of the Agulhas Current system. AWI-

CM3 displays a similar feature, with an overly strong advection of warm waters from the Indian Ocean into the Atlantic at around 40°S (Fig. A3a, b). This bias then extends northward at Atlantic Intermediate Water (AIW) depths (Fig. 9c–f). Consistent with previous studies (Gutjahr et al., 2019; von Storch et al., 2016), resolving ocean mesoscale eddies leads to a more realistic ocean vertical stratification. In line with this, the medium-resolution simulation shows a reduction in the Agulhas leakage and in the warm and salty biases near the thermocline, particularly in subtropical regions, due to improved eddy-driven mixing (Fig. 9c–f). The only exception to that is a slightly increased salinity bias in the water column above 2000 m, between 40 and 60°S , but is part of the Brazilian current rather than of the Agulhas one.

It has been shown that variations in the Agulhas system can influence thermohaline properties in the North Atlantic and potentially modulate the AMOC through the salt-advection feedback (Großelindemann et al., 2025). In our experiments, however, the low-resolution model shows only a limited AMOC response despite the stronger Agulhas leakage, whereas in the medium-resolution configuration the more realistic representation of the Agulhas system is accompanied by a stronger AMOC. A plausible explanation is that, in the LR simulation, compensatory effects between subsurface warming and salinification in the North Atlantic dampen the AMOC response. Furthermore, Zhang et al. (2025b) suggest that under present-day climate conditions the Agulhas leakage exerts only a weak influence on the AMOC, which is broadly consistent with our findings. An additional aspect worth noting is that the more realistic Agulhas representation in MR appears to be associated with a stronger sensitivity of the deep circulation to surface forcing and coupled feedbacks.

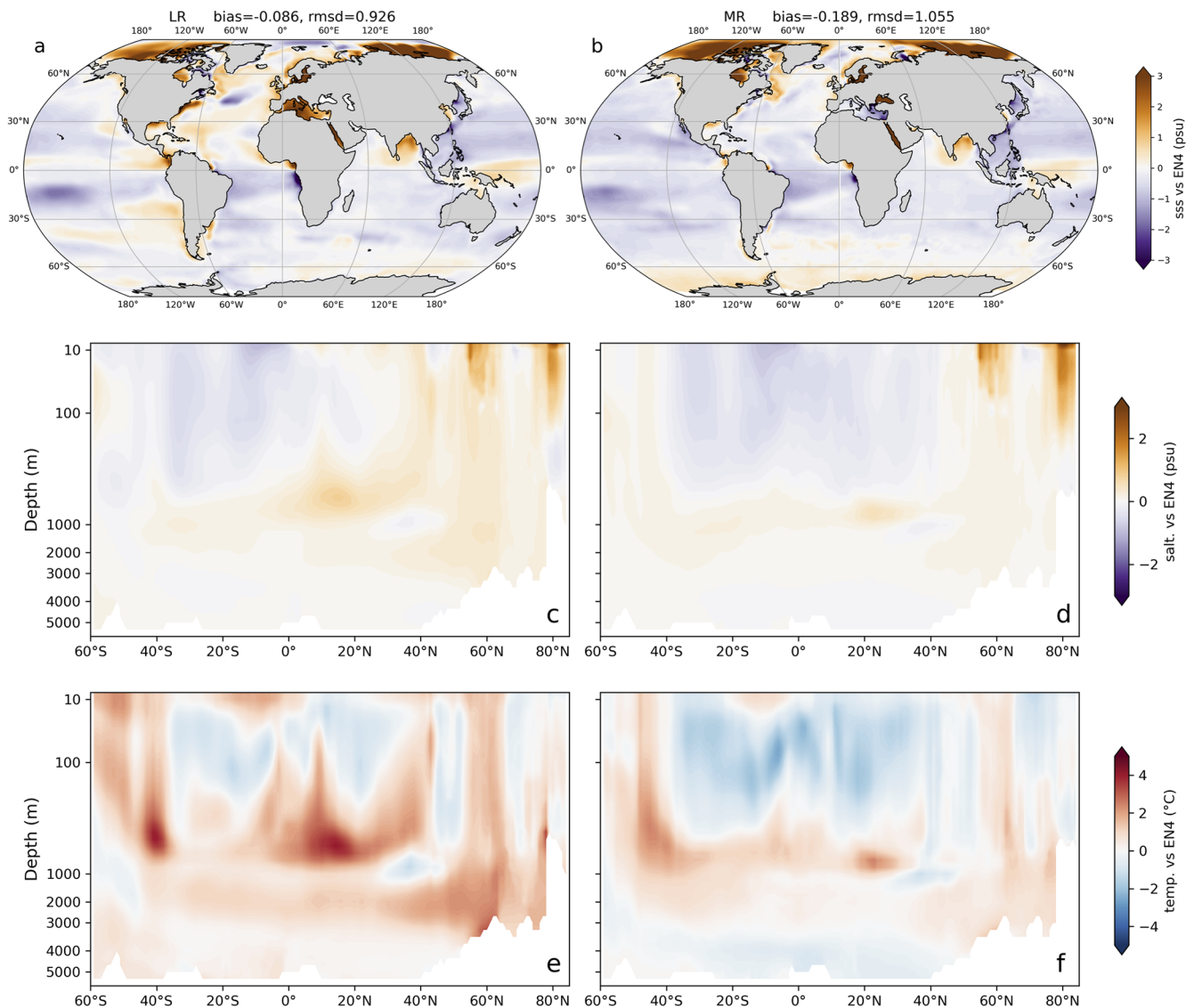


Figure 9. (a, b) Sea surface salinity mean bias and zonal mean (c, d) salinity and (e, f) temperature biases through the Atlantic Ocean and Arctic Ocean with respect to EN4. The biases, relative to the period 1979–2014, are shown for the low resolution (LR, left panels) and medium resolution (MR, right panels) simulations.

3.6 Atmosphere mean state

The AWI-CM3 low-resolution simulation shows a widespread positive bias in 500 hPa geopotential height, particularly pronounced in the Northern Hemisphere extratropics (Fig. 10a). In contrast, a negative bias dominates high southern latitudes, with particularly strong anomalies around Antarctica. In the medium-resolution simulation, the bias pattern is largely reversed in sign and the overall magnitude of the biases is substantially reduced globally (Fig. 10b).

The global zonal-mean temperature cross-sections for the low-resolution simulation (Fig. 10c) reveal a bias structure and magnitude similar to that reported by Streffing et al. (2022). In the medium-resolution the cold biases in the upper

troposphere (around 250 hPa), present in both hemispheres, persist but are reduced (Fig. 10d). An analogous, but more pronounced improvement is seen in the warm biases of the lower troposphere over polar regions as well as in the positive temperature bias in the upper stratosphere. A particularly notable change occurs in the extra-polar troposphere, where the warm bias present in the low-resolution simulation not only decreases in magnitude but also reverses sign in the medium-resolution run. The sign and vertical structure of the tropospheric temperature biases – and how they differ between the two simulations – directly explain the contrasting patterns of geopotential height bias observed in the two configurations.

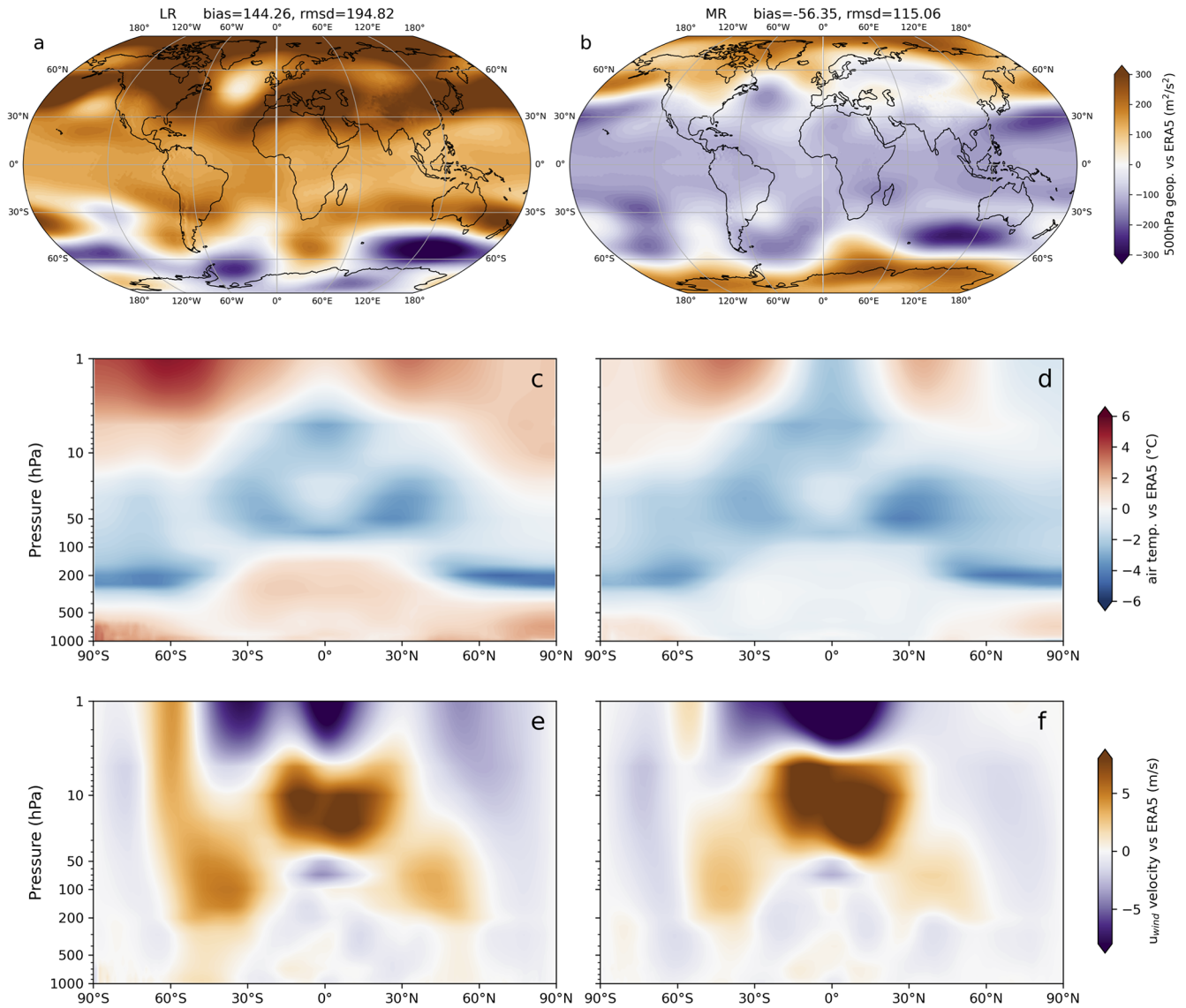


Figure 10. (a, b) 500 hPa geopotential height mean bias and global zonally averaged (c, d) air temperature and (e, f) u wind velocity biases with respect to ERA5. The biases, relative to the period 1979–2014, are shown for the low resolution (LR, left panels) and medium resolution (MR, right panels) simulations.

In CMIP6 models, tropical tropospheric warming is typically overestimated, especially between 200 and 500 hPa. This bias has been linked to surface warm biases (Mitchell et al., 2020), as well as to excessive vertical development of convection and underestimated entrainment rates in the troposphere (Keil et al., 2021). In the medium-resolution simulation presented here, surface warming is substantially reduced, and a similar reduction is seen in the tropical troposphere. This suggests that the improved representation of convective and dynamical processes – including tropical waves, mesoscale eddies, and smaller-scale convection – helps mitigate these persistent biases.

Because of the warm bias, the low-resolution simulation underestimates the equator-to-pole temperature gradient in

the troposphere (Fig. 10c), as a consequence the subtropical jet streams are displaced too far toward the poles and are simulated with excessive strength and vertical extent compared to reanalysis data (Fig. 10e). This bias is particularly pronounced in the Southern Hemisphere and helps explain the apparent decoupling between the geopotential height and tropospheric temperature biases. The overly strong, poleward-displaced, and vertically extended westerlies intensify the circumpolar trough, leading to lower geopotential heights around Antarctica despite the presence of a warm tropospheric bias.

In the medium-resolution simulation, the reduction of tropospheric temperature biases is accompanied by a substantial decrease in zonal wind biases across most latitudes

and pressure levels (Fig. 10f), indicating a more realistic thermal wind balance. In contrast, biases in equatorial stratospheric winds worsen with increased resolution. The medium-resolution simulation continues to exhibit stronger-than-observed westerlies in the upper stratosphere, indicating that AWI-CM3 still struggles to adequately reproduce the easterly phase of the Quasi-Biennial Oscillation (QBO), still a challenging aspect for climate models (Richter et al., 2020; Elsbury et al., 2021). This likely affects the representation of the Semiannual Oscillation (SAO), whose variability is strongly modulated by the QBO (Jaison et al., 2024). Accordingly, the negative wind bias in the uppermost stratosphere points to an underestimated westerly phase of the SAO.

3.7 Precipitation, cloud cover and radiation fluxes

As shown by Streffing et al. (2022), most of the positive surface temperature biases can be co-located with negative biases in cloud area fraction, and consequently with an overestimation of surface shortwave radiation (Figs. 5a, i and 11b,c). This is particularly evident in the upwelling regions off the west coasts of southern Africa and South America. Interestingly, while off the coast of South America the consistent reduction in temperature bias with increasing resolution corresponds to an improved representation of clouds and, consequently, surface radiation fluxes, the situation is different off the coast of southern Africa. There, the warm bias is reduced despite a slight deterioration in the model's representation of both clouds and radiation fluxes. A similar correspondence of patterns can be identified in the Pacific Ocean, where the temperature bias associated with the double Intertropical Convergence Zone (ITCZ) is mirrored in the bias structures of clouds, precipitation, and shortwave radiation—also as pinpointed by Streffing et al. (2022). In this low-resolution configuration of AWI-CM3, the precipitation bias is larger than the CMIP6 ensemble mean (Tian and Dong, 2020). However, increasing the model resolution leads to a considerable improvement, particularly in the tropical belt where we see an error reduction close to 20 % across all seasons (Fig. 3). A weakening of the double ITCZ bias is also visible in the medium-resolution configuration (Fig. 11a, e). Mid-to-high latitudes show a limited difference with respect to observations, in both the simulations. On a global scale a slight reduction is also observed in the shortwave radiation bias, when moving from low to medium resolution. Conversely, negative biases in clouds and longwave radiation increase in magnitude. This is primarily because the improvement in these two fields is more substantial for positive biases. Nevertheless, all the variables considered show a reduction in root mean square error (rmsd), indicating an overall improvement in model performance.

Representing cloud cover over high latitudes remains a challenging aspect for AOGCMs. In agreement with most CMIP6 models (Lauer et al., 2023), AWI-CM3 tends to overestimate cloud cover over the Southern Ocean (Fig. 11b, f).

However, this bias coexists with an overestimation of surface shortwave radiation (Fig. 11c, g), suggesting that the issue involves not only the cloud amount but also cloud properties – particularly an underrepresentation of low-level clouds or an underestimation of cloud albedo, as also noted in other CMIP6 studies (Zelinka et al., 2020; Zhao et al., 2022). Given the low surface reflectivity of the Southern Ocean, the surface albedo is particularly sensitive to cloud characteristics. Consequently, due to the strong shortwave cloud radiative cooling effect, the net cloud radiative forcing in this region is dominated by the shortwave component. This bias is likely linked to sea ice extent errors, especially during summer. For example, in the region off West Antarctica, where the positive surface shortwave radiation bias is reduced, the medium-resolution simulation shows a slight improvement in sea ice extent compared to the low-resolution configuration (Fig. A5b, d, f). Over the Arctic Basin, cloud cover overestimation has a different impact on the surface energy budget. In this region, clouds primarily act to warm the surface during winter by enhancing downward longwave radiation (Walden et al., 2017). Accordingly, AWI-CM3 shows a positive bias in surface net longwave radiation due to the excess cloud cover, while the net surface shortwave flux is underestimated (Fig. 11b–d, f–h). At lower Arctic latitudes, going from low to medium resolution, the longwave bias becomes weaker while the shortwave bias becomes more pronounced (Fig. 11c, d, g, h). This enhanced cooling effect likely contributes to the co-located overestimation of sea ice extent mentioned earlier.

Overall, it is worth noting that for all variables, the bias in the medium-resolution model configuration relative to observations is smaller in magnitude than the CMIP6 multi-model mean bias (Fig. 3). Moreover, the improvement compared to the low-resolution configuration, though modest, appears consistent across seasons and geographic regions.

3.8 The Weddell Sea atmosphere-ocean feedback

As mentioned earlier, our medium-resolution simulation shows a significant drop in September sea ice area in the Southern Ocean around year 2030, a feature not present in the low-resolution simulation (Fig. 4b). This pronounced decline, which is indicative of a strong non-linearity in the climate system, is centered in the Weddell Sea. Here approximately half of the total sea ice area is lost during the decade analyzed (Fig. 12a). In contrast, the low-resolution simulation maintains a relatively stable sea ice cover (Fig. 12i), exhibiting only slightly increased variability after mid-century and no clear decline before the last two decades of the century, consistent with the CMIP6 multi-model ensemble mean.

Prior to the onset of sea ice loss, upper ocean temperature increases steadily, though the trend differs between the two configurations: the low-resolution run shows a monotonic rise, while the medium-resolution simulation exhibits a step-

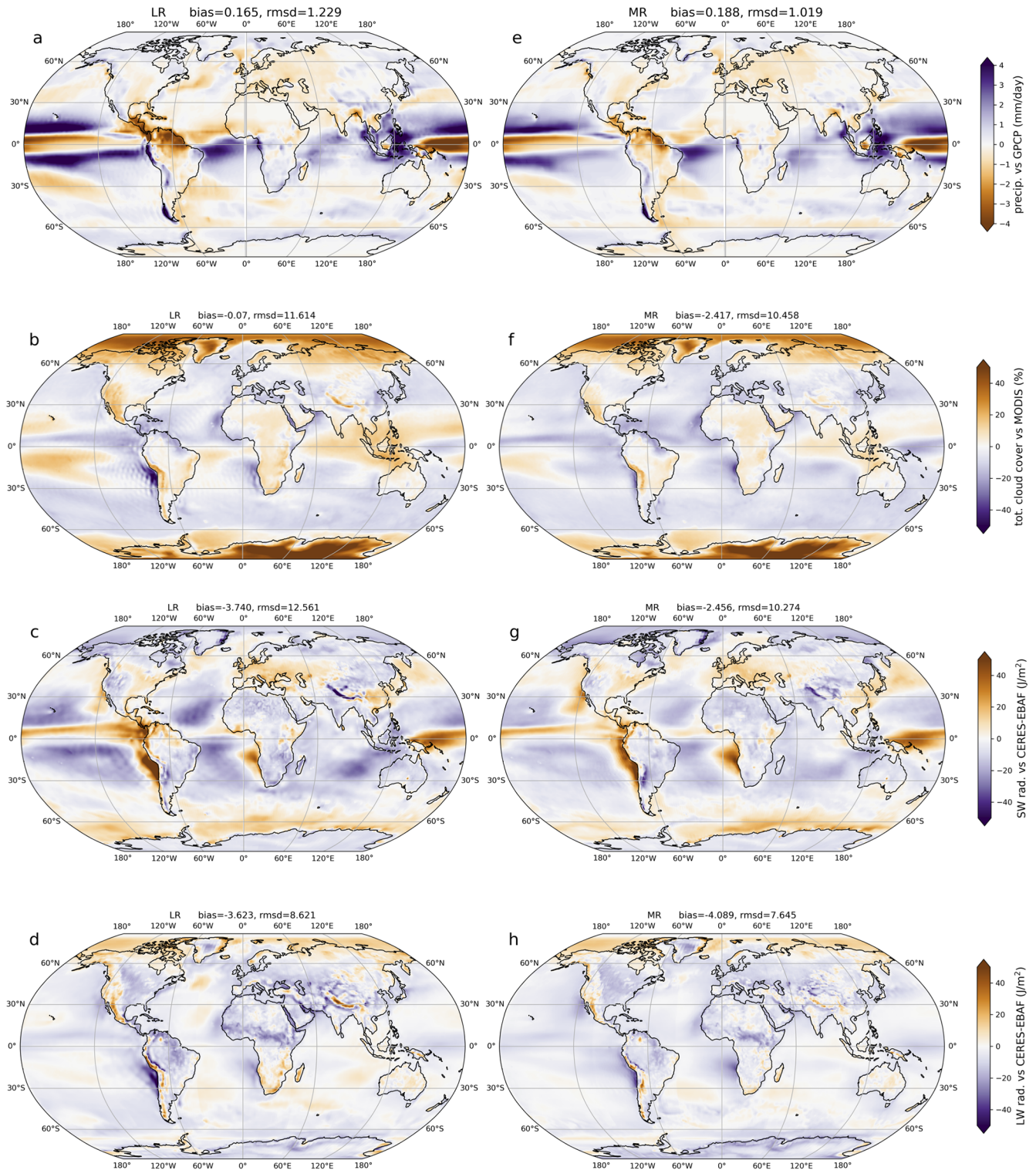


Figure 11. Annual mean (a, e) precipitation bias with respect to GPCP, (b, f) total cloud coverage bias with respect to MODIS (Platnick et al., 2015) and surface net (c, g) shortwave and (d, h) longwave radiation biases with respect to CERES (NASA/LARC/SD/ASDC, 2019; Wielicki et al., 1996). Bias and rmsd represent the area-weighted mean and root mean square difference between model and observations. Due to data availability for cloud coverage and radiation the period considered is 2000–2014, while for precipitation the period considered is 1979–2014. Results are shown for both the low resolution (LR, a–d) and medium resolution (MR, e–h) simulations.

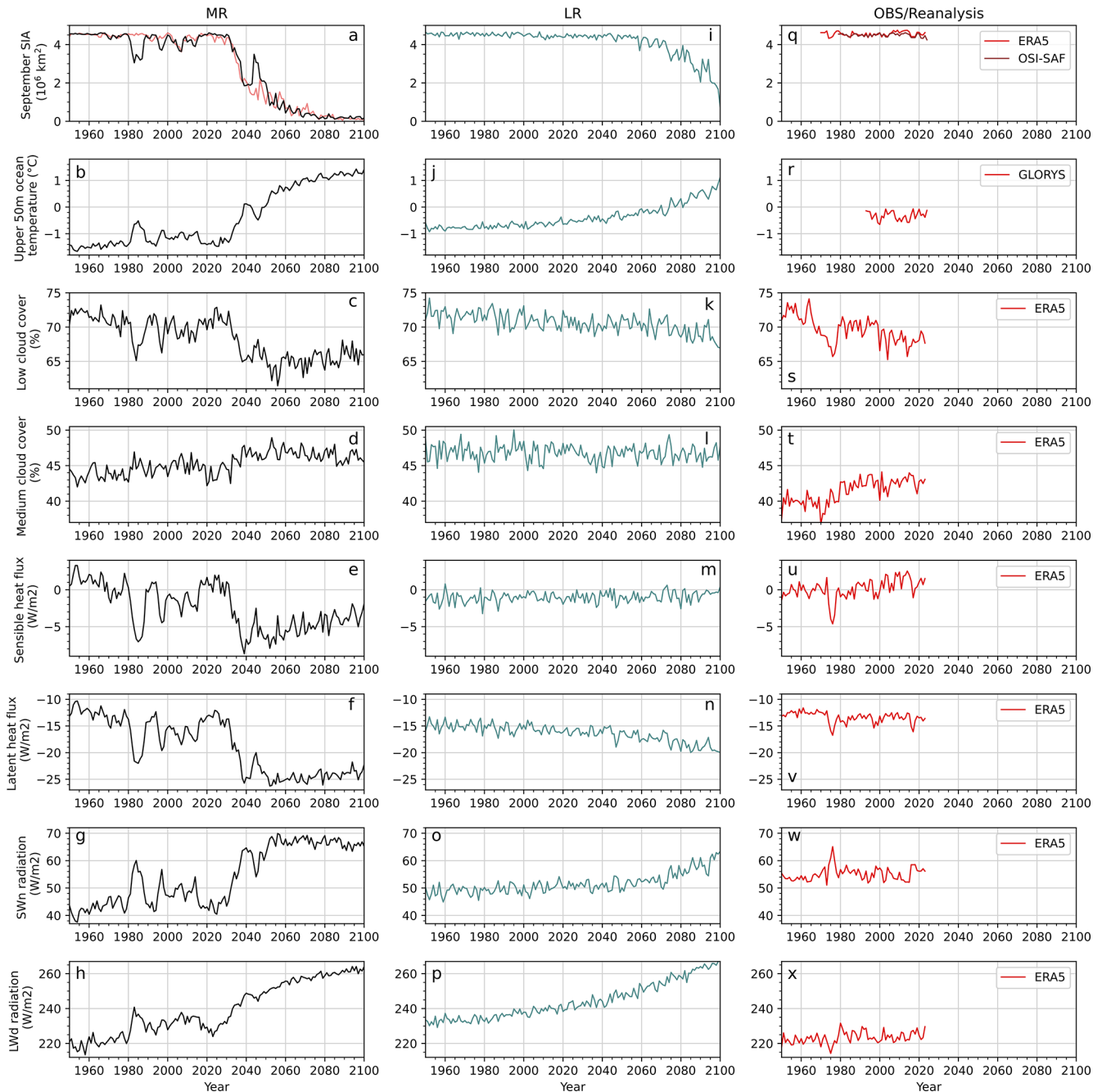


Figure 12. Time series of key variables involved in atmosphere–ocean interaction feedbacks. Panels (q)–(x) show September mean sea ice area and annual mean of upper 50 m ocean temperature, low- and mid-level cloud cover, surface sensible and latent heat fluxes, surface net shortwave radiation (SWn) and surface downward longwave radiation (LWd) over the Weddell Sea region (50° W–30° E, 60–75° S), based on observations (GLORYS, NSIDC) and reanalysis data (ERA5). Panels (a)–(p) show the same variables from the AWI-CM3 simulations at medium (TCO319-DART) and low (TCO95-CORE2) resolution. In panel (a) the light red line shows the sea ice area from the AWI-CM3 TCO319-DART simulation shown in Kim et al. (2026).

like pattern (Fig. 12b, j). As detailed by Kim et al. (2026), this marks the onset of a cascade of ocean–atmosphere feedbacks. Increased heat release from both sea ice and the ocean – via latent and sensible heat fluxes (Fig. 12e, f) – warms the lower atmosphere. Through the lapse rate and the albedo

feedbacks both tropospheric temperature and humidity shift, enhancing the surface net shortwave radiation flux (Fig. 12g) and modifying the vertical structure of the atmosphere. This affects cloud cover (Fig. 12c, d), and cloud optical depth, ultimately influencing the surface downward longwave radia-

tion flux (Fig. 12h). Figure 12 illustrates that this chain of radiative feedbacks is well captured in our medium-resolution simulation. The AWI-CM3 model version used here closely aligns with that employed by Kim et al. (2026), and both medium-resolution simulations use the same grid resolution. This allows our simulation to be considered essentially a second realization of the one presented by Kim et al. (2026), supporting the robustness of their findings. As shown in Fig. 12a, the sea ice regime shift appears in both simulations and occurs at a similar point in time. Importantly, similar feedback signatures are also identifiable in the observational and reanalysis data analyzed (Fig. 12q–x), particularly during the well-documented polynya event of 1975/1976. These processes are evident both during smaller polynya-like events that occur before the major sea ice decline around 2030, and during the large-scale decline itself. In the latter case, the magnitude of the variability is markedly stronger, and unlike the earlier events, no recovery of sea ice extent is observed in the Weddell Sea region thereafter. In contrast, the low-resolution simulation fails to reproduce this sequence of feedbacks. Across all examined fields, it shows a persistently monotonic and stable behavior, lacking the coupled dynamical and radiative variability seen in both the medium-resolution run and observational data.

It can be noted that, when comparing the timeseries, the medium-resolution simulation tends to overestimate the amplitude of variability relative to reanalysis data. However, studies have shown that ocean reanalyses often disagree in their representation of seasonal and interannual variability (Hobbs et al., 2020; Nakayama et al., 2024), particularly on regional scales, and that Southern Ocean variability can exceed what is captured by existing reanalyses or sparse observations (Shi et al., 2021; Auger et al., 2021; Nakayama et al., 2024). In conclusion, these results support the suggestion of a potential sudden, accelerated Antarctic sea-ice decline in the coming decades, but also show the evidence for a non-linear sea ice sensitivity to greenhouse warming, which can contribute to explain the apparent mismatch between the current generation of climate models and observations.

4 Conclusions

This study presents a comprehensive evaluation of the AWI-CM3 coupled climate model in both low-resolution (LR) and medium-resolution (MR) configurations, focusing on their ability to simulate key aspects of Earth's climate across a range of variables, time periods, and spatial domains. Through an experimental setup in which both configurations followed the CMIP6 HighResMIP protocol and were run over the same historical and future periods, we were able to isolate and assess the impact of horizontal resolution on the model's performance. The results reveal consistent and substantial improvements in the medium-resolution simulation, which achieved an average reduction of $\sim 28\%$ in climato-

logical errors relative to the CMIP6 multi-model mean and showed superior agreement with observations across most variables, regions, and seasons.

The benefits of higher resolution are particularly evident in regions where mesoscale and submesoscale processes play a crucial role, such as the polar oceans. In the Southern Ocean, the MR configuration substantially improves the representation of sea ice dynamics, mixed-layer depth variability, and deep convection patterns, leading to more realistic simulations of polynya formation and regional heat fluxes.

The medium-resolution model also exhibits a marked regime shift in Antarctic sea ice extent in the Weddell Sea (Kim et al., 2026), triggered by ocean-atmosphere feedbacks, which is absent in the LR simulation and most CMIP6 models. This underscores how higher-resolution models, with eddy-resolving oceans and improved atmospheric representation, can capture emergent behaviors and abrupt climate transitions that may be overlooked in lower-resolution configurations (Roberts et al., 2025). Such capability is of increasing concern in light of the rapid and potentially irreversible environmental changes observed in recent years in Antarctica (Abram et al., 2025). This is particularly important also given that it remains uncertain whether upcoming AI-based climate models will be able to accurately reproduce feedback processes or identify potential sudden regime shifts.

In the Northern Hemisphere, the MR simulation offers improved spatial and temporal representations of sea ice variability, although some biases – such as the overestimated September sea ice extent in mid-latitudes – persist. Across the tropics and mid-latitudes, increased resolution leads to a more accurate depiction of surface temperature trends, precipitation patterns, and ocean heat content, although improvements are regionally dependent and sometimes accompanied by enhanced variability. In terms of atmospheric dynamics, the MR model reduces longstanding biases in tropospheric temperature and zonal wind fields, leading to a more realistic thermal wind balance and an improved simulation of the subtropical jet streams. While stratospheric biases, particularly related to the Quasi-Biennial Oscillation (QBO), remain, overall fidelity in large-scale circulation features is markedly enhanced.

One of the more compelling findings of this study is the improvement in ocean circulation patterns and volume transports through major straits and channels in the MR simulation. Enhanced resolution enables the model to better resolve mesoscale eddies and vertical mixing processes, leading to more realistic simulations of features such as the Atlantic Meridional Overturning Circulation (AMOC), Arctic outflows, and Antarctic Bottom Water (AABW) formation. These enhancements contribute to a more accurate portrayal of heat and salt transport across ocean basins, as well as feedbacks between sea ice, ocean dynamics, and the atmosphere.

Medium-resolution simulations offer several advantages, including a more detailed representation of interannual to

decadal variability and feedback processes, which can reveal different potential trajectories for the climate system and provide a richer picture of possible climate outcomes. This enhanced variability, at the same time, can make it more challenging to attribute forced trends and separate signal from internal noise. As noted earlier, only a single ensemble member was run and analyzed. The higher computational cost makes it more challenging – albeit not impossible – to generate large ensembles, which are important for robust uncertainty quantification and for assessing the statistical significance of model responses. Consequently, relying on a single simulation realization inherently constrains the ability to fully sample internal climate variability or to disentangle stochastic fluctuations from externally forced changes. Nonetheless, even a single MR simulation allows us to formulate hypotheses and gain valuable insights into model performance, emergent behavior, and potential climate responses. It highlights the strengths of higher resolution while identifying areas where ensemble approaches could further strengthen confidence in projections.

In conclusion, our findings emphasize the importance of increased horizontal resolution in improving the fidelity of coupled climate simulations, particularly in regions where small-scale processes and feedback mechanisms are critical. The MR configuration of AWI-CM3 demonstrates a clear step forward in reproducing observed climatology, resolving nonlinear climate dynamics, and capturing potential tipping points in the Earth system. As the climate modeling community advances toward CMIP7 and future generations of Earth system models, our results advocate for the integration of at least medium-resolution (or higher) modeling as a complementary strategy to traditional multi-model ensembles. These efforts, paired with rigorous observational benchmarking and continued refinement of parameterizations, will be essential in reducing uncertainties in climate projections and improving the credibility of climate change assessments.

Appendix A: Additional Figures

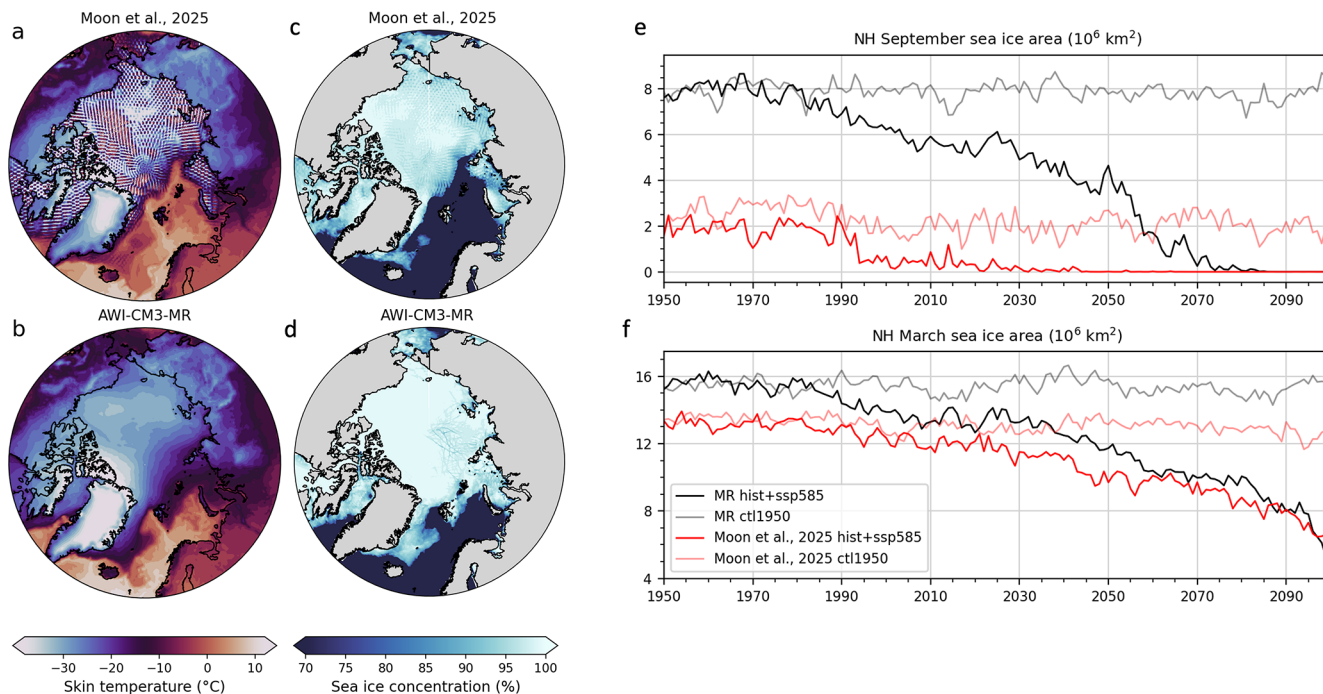


Figure A1. Comparison of skin temperature and sea ice concentration for the area north of 60° S, as well as total September/March sea ice area, between the TCo319L137-DART simulation described in Moon et al. (2025) and the AWI-CM3-MR simulation evaluated in this study. The aim is to highlight improvements resulting from the modified interpolation method for sea ice heat fluxes. Skin temperature is shown as the monthly mean for March 2000, while sea ice concentration is shown as the daily mean for 1 March 2000.

Table A1. List of CMIP variables used in Figs. 3 and A2 given together with their corresponding standard names and units.

CMIP standard name	Extended variable name (unit)
siconc	Sea ice concentration (%)
tas	Near-surface air temperature (K)
clt	Total cloud cover (%)
pr	Precipitation flux ($\text{kg m}^{-2} \text{s}^{-1}$)
rlut	Upward longwave radiation at top of atmosphere (W m^{-2})
uas	Eastward near-surface wind (m s^{-1})
vas	Northward near-surface wind (m s^{-1})
300 hPa ua	Zonal wind at 300 hPa (m s^{-1})
500 hPa zg	Geopotential height at 500 hPa (m)
zos	Sea surface height above geoid (m)
tos	Sea surface temperature ($^{\circ}\text{C}$)
m1otst	Ocean mixed layer thickness (m)
10 m thetao	Ocean potential temperature at 10 m depth ($^{\circ}\text{C}$)
100 m thetao	Ocean potential temperature at 100 m depth ($^{\circ}\text{C}$)
1000 m thetao	Ocean potential temperature at 1000 m depth ($^{\circ}\text{C}$)
10 m so	Ocean salinity at 10 m depth (1×10^{-3})
100 m so	Ocean salinity at 100 m depth (1×10^{-3})
1000 m so	Ocean salinity at 1000 m depth (1×10^{-3})

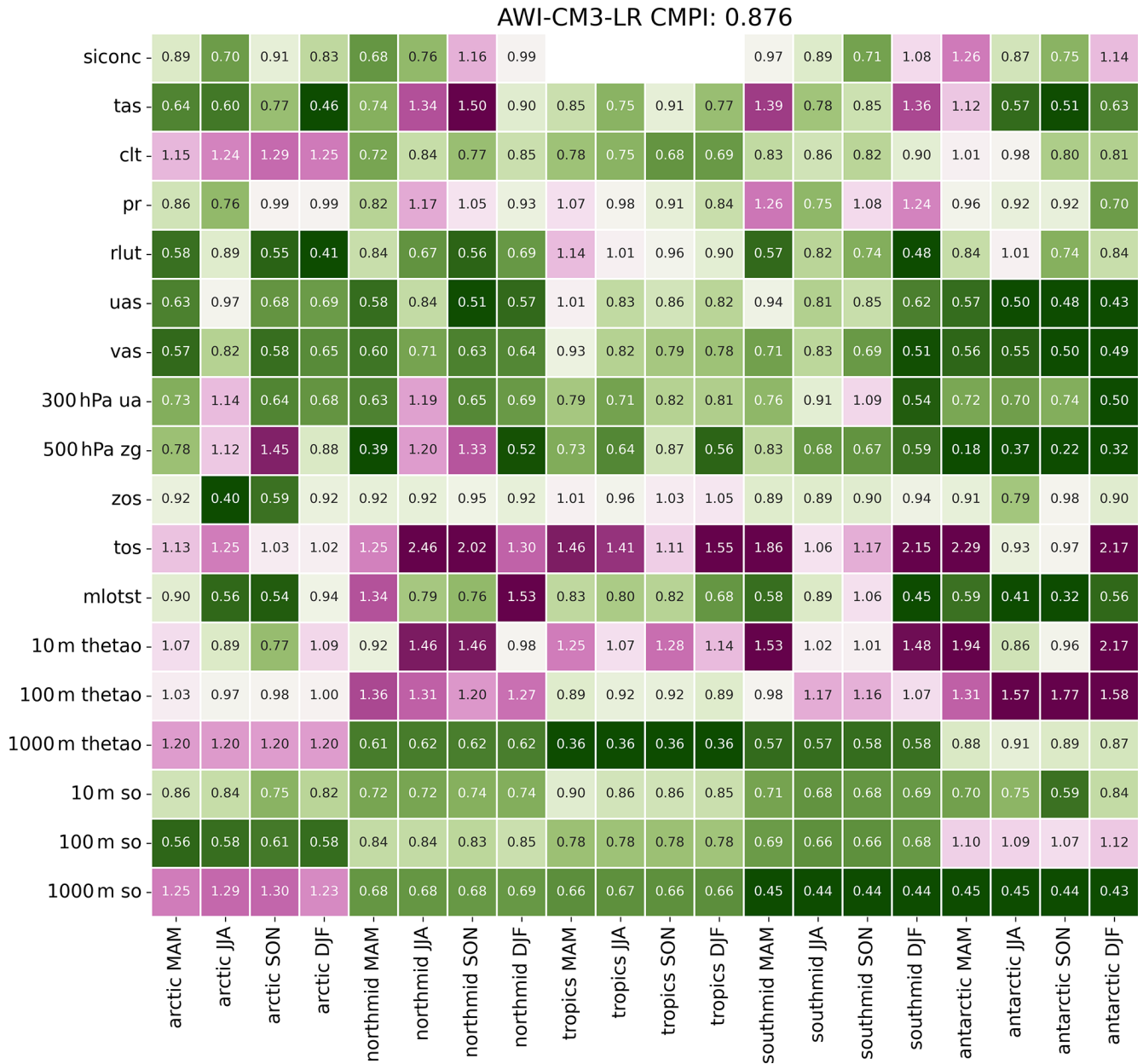


Figure A2. Performance indices that give the absolute error in 1989–2014 AWI-CM3 historic climatology as a fraction of the absolute error averaged over CMIP6 models (Steffing et al., 2025; Reichler and Kim, 2008). Values below (above) 1 correspond to below (above) CMIP6 average biases. The indices here reported are computed for the low (AWI-CM3-LR) resolution simulation. The Arctic and Antarctic regions are the areas north of 60° N and south of 60° S, respectively. The mid-latitudes span between 30 and 60° in both hemispheres, while the tropical region lies between 30° N and 30° S.

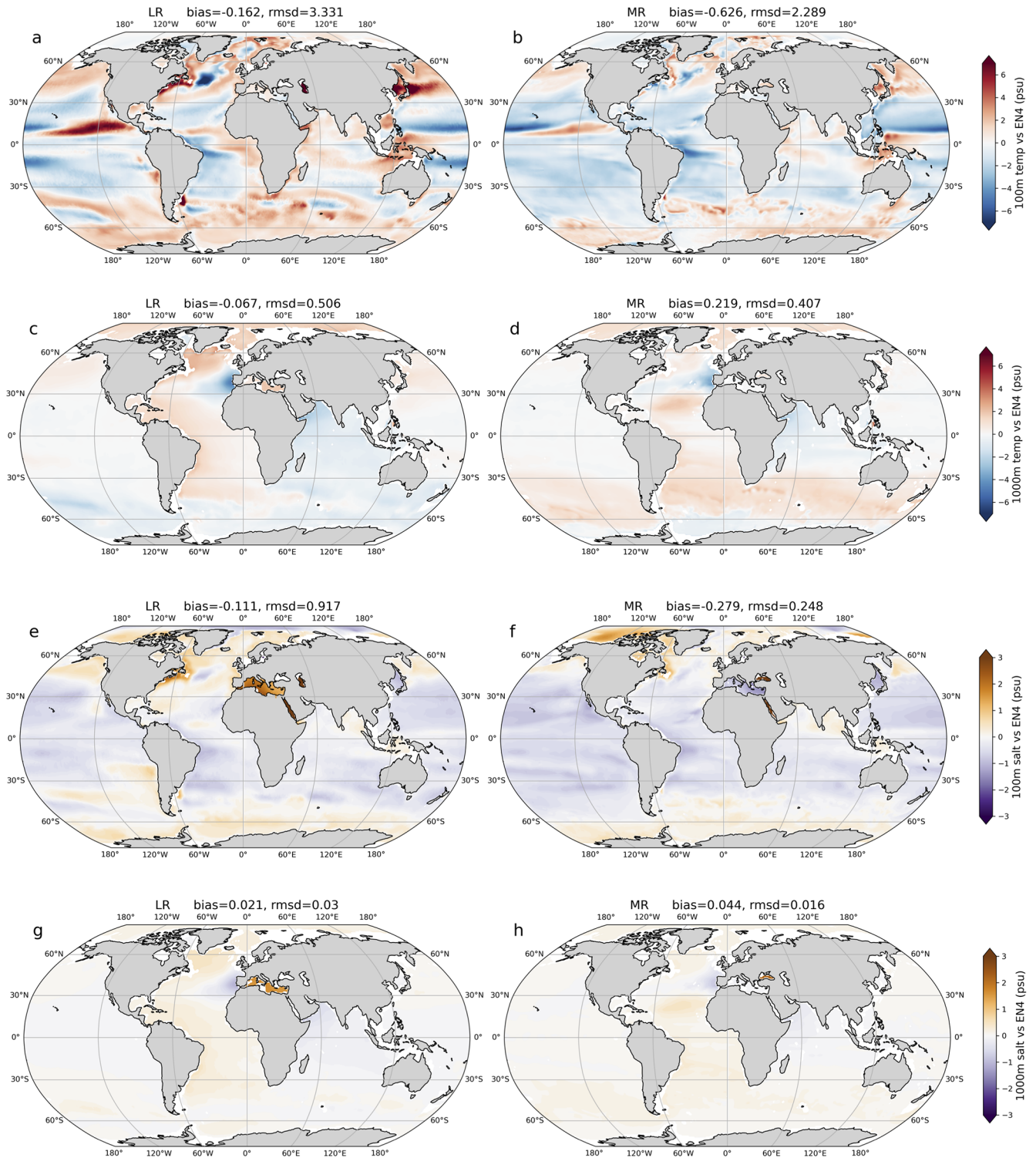


Figure A3. Ocean (a–d) temperature and (e–h) salinity biases with respect to EN4 dataset at 100 and 1000 m depth. The biases, relative to the period 1979–2014, are shown for the low resolution (LR, left panels) and medium resolution (MR, right panels) simulations.

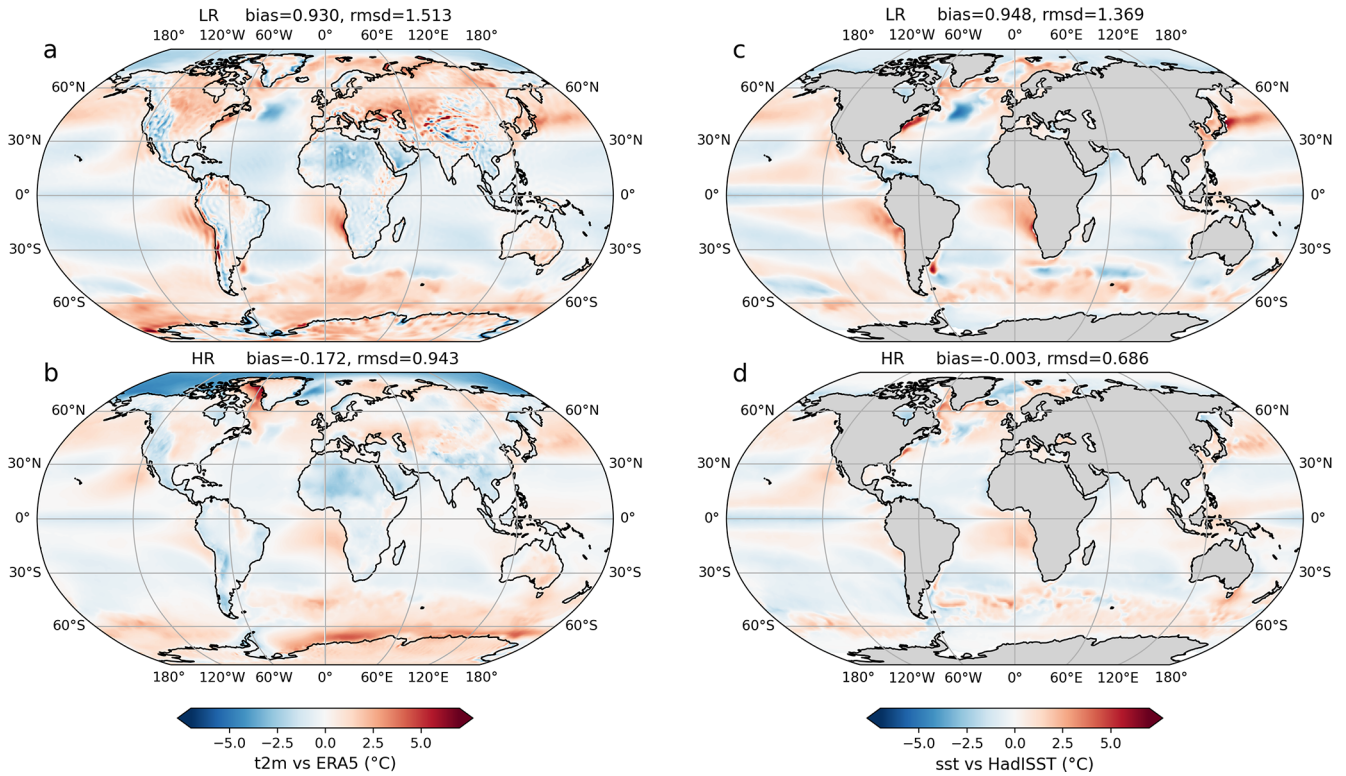


Figure A4. Annual mean 2 m air temperature (a, b) and sea surface temperature (c, d) biases (global mean bias removed) for the period 1979–2014 relative to ERA5 and HadISST, respectively. Results are shown for both the low-resolution (LR) and medium-resolution (MR) experiments.

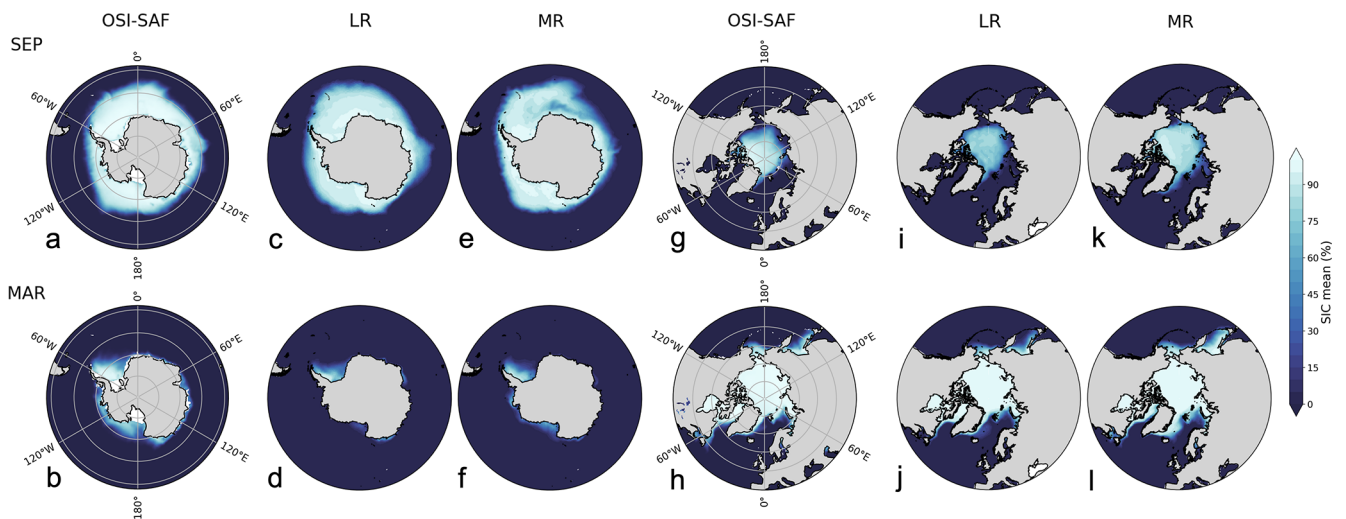


Figure A5. Means of March and September sea ice concentration from 1979 to 2014 are shown for both hemispheres, based on OSI-SAF observations (a, b, g, h) and simulations from the low-resolution (LR, c, d, i, j) and medium-resolution (MR, e, f, k, l) model configurations.

Code and data availability. ERA5 monthly averaged reanalysis data (on single and pressure levels) are publicly available from the European Centre for Medium-Range Weather Forecasts (ECMWF) Climate Data Store (CDS) at <https://cds.climate.copernicus.eu/> (last access: 30 May 2025). GLORYS ocean reanalysis products were obtained from the Copernicus Marine Environment Monitoring Service (CMEMS) and can be accessed at <https://doi.org/10.48670/moi-00021> (Mercator Ocean International, 2018). HadISST1 and EN4 datasets are produced by the UK Met Office Hadley Centre and can be retrieved from their website <https://www.metoffice.gov.uk/hadobs/> (last access: 20 May 2025). Sea ice data are part of the NSIDC-0051, produced by the National Snow and Ice Data Center (NSIDC), and OSI-450-a1, produced by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) through the OSI SAF program, datasets that can be accessed and downloaded from <https://doi.org/10.5067/MPYG15WAA4WX> (DiGirolamo et al., 2017) and <ftp://osisaf.met.no/> (last access: 26 May 2025). RAPID array data can be accessed through the RAPID/MOCHA/WBTS project website: <https://www.rapid.ac.uk/rapidmoc/> (last access: 25 April 2025). Clouds observations are part of the MOD08-M3 dataset, produced by NASA's MODIS Science Team, can be retrieved from the LAADS DAAC Portal: <https://ladsweb.modaps.eosdis.nasa.gov/> (last access: 9 June 2025). Surface fluxes belong to the Clouds and the Earth's Radiant Energy System (CERES) dataset, produced by NASA Langley Research Center (LaRC) Atmospheric Science Data Center (ASDC), are available from the ASDC Data portal <https://asdc.larc.nasa.gov/project/CERES> (last access: 9 June 2025).

The total raw model data generated during the present study are currently archived at the German Climate Computing Center (DKRZ). Due to the very large data volume, the complete dataset cannot be made publicly available for direct download. However, all data are preserved under long-term storage at DKRZ and can be accessed through collaborative arrangements or data services subject to DKRZ access policies. The subset of model output used in the analyses presented here, is available in the Zenodo database under <https://doi.org/10.5281/zenodo.17377570> (Zapponini, 2025a) for the LR and under <https://doi.org/10.5281/zenodo.17397357> (Zapponini, 2025b) for the MR. Jupyter Notebooks to reproduce the figures are available on Zenodo at <https://doi.org/10.5281/zenodo.17996211> (Zapponini, 2025c).

The OpenIFS model is available at <http://www.ecmwf.int/en/research/projects/openifs> (last access: 17 October 2025) via an OpenIFS licence. The specific configuration used in this study is available from the DKRZ GitLab repository <https://gitlab.dkrz.de/ec-earth/oifs-43r3/> (last access: 17 October 2025) and corresponds to the Git tag `Zapponini_et_al_2025` on the branch `origin/aerosol_scaling_for_v3.1.1_martina`. The exact configuration of the FESOM2 ocean model used for this study is available on Zenodo at <https://doi.org/10.5281/zenodo.17961041> (Scholz et al., 2025). It corresponds to the Git tag `Zapponini_et_al_2025` on the branch `origin/mzapponi_refactoring4awicm3.1` of the FESOM2 GitHub repository.

Author contributions. MZ wrote the initial manuscript. MZ conducted the experiment with the help of TS and JS, MZ performed all the analysis of the results. All the authors contributed to the scientific discussion and reviewed the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. All the analyses were performed at the German Climate Computing Center (DKRZ), in the framework of the project ab0995, on the Levante supercomputer.

Financial support. JS received funding from the European Union's Horizon Europe research and innovation programme through the LIQUIDICE project (grant no. 101184962), and from the Federal Ministry of Research, Technology and Space of Germany within the CAP7 project (grant no. 01LP2401G). LAR was supported by the Initiative and Networking Fund of the Helmholtz Association (grant no. VH-NG-20-10). TJ received the financial support by the Federal Ministry of Research, Technology and Space of Germany in the framework of codeBetter project (grant no. 01LK2202A) and EERIE project (grant no. 101081383) funded by the European Union. The work presented in this paper has been produced in the context of the European Union's Destination Earth Initiative and relates to tasks entrusted by the European Union to the European Centre for Medium-Range Weather Forecasts implementing part of this Initiative with funding by the European Union.

The article processing charges for this open-access publication were covered by the Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung.

Review statement. This paper was edited by Paul Ullrich and reviewed by two anonymous referees.

References

Abram, N. J., Purich, A., England, M. H., McCormack, F. S., Strugnell, J. M., Bergstrom, D. M., Vance, T. R., Stål, T., Wie-

- necke, B., Heil, P., Doddridge, E. D., Sallée, J.-B., Williams, T. J., Reading, A. M., Mackintosh, A., Reese, R., Winkelmann, R., Klose, A. K., Boyd, P. W., Chown, S. L., and Robinson, S. A.: Emerging evidence of abrupt changes in the Antarctic environment, *Nature*, 644, 621–633, <https://doi.org/10.1038/s41586-025-09349-5>, 2025.
- Athanase, M., Köhler, R., Heuzé, C., Lévine, X., and Williams, R.: The Arctic Beaufort Gyre in CMIP6 Models: Present and Future, *J. Geophys. Res.-Oceans*, 130, e2024JC021873, <https://doi.org/10.1029/2024JC021873>, 2025.
- Auger, M., Morrow, R., Kestenare, E., Sallée, J.-B., and Cowley, R.: Southern Ocean in-situ temperature trends over 25 years emerge from interannual variability, *Nat. Commun.*, 12, 514, <https://doi.org/10.1038/s41467-020-20781-1>, 2021.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, *J. Hydrometeorol.*, 10, 623–643, <https://doi.org/10.1175/2008JHM1068.1>, 2009.
- Beadling, R. L., Russell, J., Stouffer, R., Mazloff, M., Talley, L., Goodman, P., Sallée, J.-B., Hewitt, H., Hyder, P., and Pandde, A.: Representation of Southern Ocean properties across coupled model intercomparison project generations: CMIP3 to CMIP6, *J. Climate*, 33, 6555–6581, <https://doi.org/10.1175/JCLI-D-19-0970.1>, 2020.
- Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Soci, C., Villaume, S., Bidlot, J.-B., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N.: The ERA5 global reanalysis: Preliminary extension to 1950, *Q. J. Roy. Meteor. Soc.*, 147, 4186–4227, <https://doi.org/10.1002/qj.4174>, 2021.
- Beszczynska-Möller, A., Von Appen, W., and Fahrbach, E.: Physical oceanography and current meter data from moorings F1-F14 and F15/F16 in the Fram Strait, 1997–2012, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.150016>, 2015.
- Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Mehl, G., Predoi, V., Roberts, M., and Eyring, V.: Quantifying progress across different CMIP phases with the ESMValTool, *J. Geophys. Res.-Atmos.*, 125, e2019JD032321, <https://doi.org/10.1029/2019JD032321>, 2020.
- Casagrande, F., Stachelski, L., and de Souza, R. B.: Assessment of Antarctic sea ice area and concentration in Coupled Model Intercomparison Project Phase 5 and Phase 6 models, *Int. J. Climatol.*, 43, 1314–1332, <https://doi.org/10.1002/joc.7916>, 2023.
- Cesana, G. V., Khadir, T., and Chepfer, Hand Chiriaco, M.: Southern Ocean solar reflection biases in CMIP6 models linked to cloud phase and vertical structure representations, *Geophys. Res. Lett.*, 49, e2022GL099777, <https://doi.org/10.1029/2022GL099777>, 2022.
- Cesana, G. V., Roach, L. A., and Blanchard-Wrigglesworth, E.: Clouds are crucial to capture Antarctic sea ice variability, *Geophys. Res. Lett.*, 52, e2024GL113322, <https://doi.org/10.1029/2024GL113322>, 2025.
- Chen, D., Rojas, M., Samset, B. H., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S. H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S. K., Plattner, G.-K., and Tréguier, A.-M.: Framing, Context, and Methods, in: *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 147–286, <https://doi.org/10.1017/9781009157896.003>, 2021.
- Craig, A., Valcke, S., and Coquart, L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, *Geosci. Model Dev.*, 10, 3297–3308, <https://doi.org/10.5194/gmd-10-3297-2017>, 2017.
- Curry, B., Lee, C. M., Petrie, B., Moritz, R. E., and Kwok, R.: Multiyear volume, liquid freshwater, and sea ice transports through Davis Strait, 2004–10, *J. Phys. Oceanogr.*, 44, 1244–1266, <https://doi.org/10.1175/JPO-D-13-0177.1>, 2014.
- Danabasoglu, G., McWilliams, J. C., and Large, W. G.: Approach to equilibrium in accelerated global oceanic models, *J. Climate*, 1092–1110, [https://doi.org/10.1175/1520-0442\(1996\)009<1092:ATELAG>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1092:ATELAG>2.0.CO;2), 1996.
- Danek, C., Scholz, P., and Lohmann, G.: Effects of high resolution and spinup time on modeled North Atlantic circulation, *J. Phys. Oceanogr.*, 49, 1159–1181, <https://doi.org/10.1175/JPO-D-18-0141.1>, 2019.
- Danilov, S., Wang, Q., Timmermann, R., Iakovlev, N., Sidorenko, D., Kimmritz, M., Jung, T., and Schröter, J.: Finite-Element Sea Ice Model (FESIM), version 2, *Geosci. Model Dev.*, 8, 1747–1761, <https://doi.org/10.5194/gmd-8-1747-2015>, 2015.
- Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volume Sea ice–Ocean Model (FESOM2), *Geosci. Model Dev.*, 10, 765–789, <https://doi.org/10.5194/gmd-10-765-2017>, 2017.
- De Boer, A., Gavilan P-A, E., Stevens, D., Chafik, L., Hutchinson, D., Zhang, Q., Sime, L., and Willmott, A.: Interconnectivity between volume transports through Arctic straits, *J. Geophys. Res.-Oceans*, 123, 8714–8729, <https://doi.org/10.1029/2018JC014320>, 2018.
- DiGirolamo, N., Parkinson, C. L., Cavalieri, D. J., Gloersen, P., and Zwally, H. J.: Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data (NSIDC-0051, Version 2), NASA National Snow and Ice Data Center Distributed Active Archive Center [data set], <https://doi.org/10.5067/MPYG15WAA4WX>, 2017.
- Doblas-Reyes, F. J., Kontkanen, J., Sandu, I., Acosta, M., Al Turjman, M. H., Alsina-Ferrer, I., Andrés-Martínez, M., Anerdi, C., Arriola, L., Axness, M., Battle Martín, M., Bauer, P., Becker, T., Beltrán, D., Beyer, S., Bockelmann, H., Bretonnière, P.-A., Cabaniols, S., Caprioli, S., Castrillo, M., Chandrasekar, A., Cheedela, S., Correal, V., Danovaro, E., Davini, P., Enkovaara, J., Frauen, C., Früh, B., Gaya Àvila, A., Ghinassi, P., Ghosh, R., Ghosh, S., González, I., Grayson, K., Griffith, M., Hadade, I., Haine, C., Hartick, C., Haus, U.-U., Hearne, S., Järvinen, H., Jiménez, B., John, A., Juchem, M., Jung, T., Kegel, J., Kelbling, M., Keller, K., Kinoshita, B., Kiszler, T., Klocke, D., Kluft, L., Koldunov, N., Kölling, T., Kolstela, J., Kornblueh, L., Kosukhin, S., Lacima-Nadolnik, A., Leal Rojas, J. J., Lehtiranta, J., Lunttila, T., Luoma, A., Manninen, P., Medvedev, A., Milinski, S., Mohammed, A., Müller, S., Naryanappa, D., Nazarova, N., Niemelä, S., Niraula, B., Nortamo, H., Nummelin, A., Nurisso,

- M., Ortega, P., Paronuzzi, S., Pedruzo-Bagazgoitia, X., Pelletier, C., Peña, C., Polade, S., Pradhan, H. K., Quintanilla, R., Quintino, T., Rackow, T., Räisänen, J., Rajput, M. M., Redler, R., Reuter, B., Rocha Monteiro, N., Roura-Adserias, F., Ruppert, S., Sayed, S., Schnur, R., Sharma, T., Sidorenko, D., Sievi-Korte, O., Soret, A., Steger, C., Stevens, B., Streffing, J., Sunny, J., Tenorio, L., Thober, S., Tigerstedt, U., Tinto, O., Tonttila, J., Tuomenvirta, H., Tuppi, L., Van Thielen, G., Vitali, E., von Hardenberg, J., Wagner, I., Wedi, N., Wehner, J., Willner, S., Yepes-Arbós, X., Ziemer, F., and Zimmermann, J.: The Destination Earth digital twin for climate change adaptation, *Geosci. Model Dev.*, 19, 2821–2848, <https://doi.org/10.5194/gmd-19-2821-2026>, 2026.
- Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P., and Chereskin, T. K.: Mean antarctic circumpolar current transport measured in drake passage, *Geophys. Res. Lett.*, 43, 11–760, <https://doi.org/10.1002/2016GL070319>, 2016.
- ECMWF: IFS Documentation CY43R3 – Part III: Dynamics and numerical procedures, ECMWF, <https://doi.org/10.21957/817miod5m>, 2017a.
- ECMWF: IFS Documentation CY43R3 – Part IV: Physical processes, ECMWF, <https://doi.org/10.21957/efyk72kl>, 2017b.
- ECMWF: IFS Documentation CY43R3 – Part VII: ECMWF wave model, ECMWF, <https://doi.org/10.21957/mxz9z1gb>, 2017c.
- Elsbury, D., Peings, Y., and Magnusdottir, G.: CMIP6 models underestimate the Holton-Tan effect, *Geophys. Res. Lett.*, 48, e2021GL094083, <https://doi.org/10.1029/2021GL094083>, 2021.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 423–552, <https://doi.org/10.1017/9781009157896.005>, 2021.
- Gent, P. R. and McWilliams, J. C.: Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr.*, 20, 150–155, [https://doi.org/10.1175/1520-0485\(1990\)020<0150:IMIOCM>2.0.CO;2](https://doi.org/10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2), 1990.
- Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *J. Geophys. Res.-Oceans*, 118, 6704–6716, <https://doi.org/10.1002/2013JC009067>, 2013.
- Gordon, A. L., Sprintall, J., Van Aken, H. M., Susanto, D., Wijffels, S., Molcard, R., Ffield, A., Pranowo, W., and Wirasantosa, S.: The Indonesian throughflow during 2004–2006 as observed by the INSTANT program, *Dynam. Atmos. Oceans*, 50, 115–128, <https://doi.org/10.1016/j.dynatmoce.2009.12.002>, 2010.
- Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., Dunne, J. P., Goddard, P., Morrison, A. K., Rosati, A., Wittenberg, A. T., Yin, J., and Zhang, R.: Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models, *J. Climate*, 28, 952–977, <https://doi.org/10.1175/JCLI-D-14-00353.1>, 2015.
- Großelindemann, H., Castruccio, F. S., Danabasoglu, G., and Biasi, A.: Long-term variability and trends in the Agulhas Leakage and its impacts on the global overturning, *Ocean Sci.*, 21, 93–112, <https://doi.org/10.5194/os-21-93-2025>, 2025.
- Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., Haak, H., and Stössel, A.: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), *Geosci. Model Dev.*, 12, 3241–3281, <https://doi.org/10.5194/gmd-12-3241-2019>, 2019.
- Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, *Geosci. Model Dev.*, 9, 4185–4208, <https://doi.org/10.5194/gmd-9-4185-2016>, 2016.
- Hallberg, R.: Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects, *Ocean Model.*, 72, 92–103, <https://doi.org/10.1016/j.ocemod.2013.08.007>, 2013.
- Hellmuth, F., Carlsen, T., Daloz, A. S., David, R. O., Che, H., and Storelvmo, T.: Evaluation of biases in mid-to-high-latitude surface snowfall and cloud phase in ERA5 and CMIP6 using satellite observations, *Atmos. Chem. Phys.*, 25, 1353–1383, <https://doi.org/10.5194/acp-25-1353-2025>, 2025.
- Heuzé, C. and Liu, H.: No emergence of deep convection in the Arctic Ocean across CMIP6 models, *Geophys. Res. Lett.*, 51, e2023GL106499, <https://doi.org/10.1029/2023GL106499>, 2024.
- Heuzé, C., Heywood, K. J., Stevens, D. P., and Ridley, J. K.: Southern Ocean bottom water characteristics in CMIP5 models, *Geophys. Res. Lett.*, 40, 1409–1414, <https://doi.org/10.1002/grl.50287>, 2013.
- Heuzé, C., Zanolini, H., Karam, S., and Muilwijk, M.: The deep arctic ocean and fram strait in cmip6 models, *J. Climate*, 36, 2551–2584, 2023.
- Hewitt, H., Fox-Kemper, B., Pearson, B., Roberts, M., and Klocke, D.: The small scales of the ocean may hold the key to surprises, *Nat. Clim. Change*, 12, 496–499, <https://doi.org/10.1038/s41558-022-01386-6>, 2022.
- Hobbs, W. R., Klekociuk, A. R., and Pan, Y.: Validation of reanalysis Southern Ocean atmosphere trends using sea ice data, *Atmos. Chem. Phys.*, 20, 14757–14768, <https://doi.org/10.5194/acp-20-14757-2020>, 2020.
- Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behraves, M., Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datsieris, G., Esch, M., George, G., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornbluh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T.,

- Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemer, F., Marotzke, J., and Stevens, B.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, *Geosci. Model Dev.*, 16, 779–811, <https://doi.org/10.5194/gmd-16-779-2023>, 2023.
- Jaison, A. M., Gray, L. J., Osprey, S. M., Knight, J. R., and Andrews, M. B.: Role of the quasi-biennial oscillation in alleviating biases in the semi-annual oscillation, *Weather Clim. Dynam.*, 5, 1489–1504, <https://doi.org/10.5194/wcd-5-1489-2024>, 2024.
- Johnson, G. C., Lyman, J. M., and Loeb, N. G.: Improving estimates of Earth's energy imbalance, *Nat. Clim. Change*, 6, 639–640, <https://doi.org/10.1038/nclimate3043>, 2016.
- Keil, P., Schmidt, H., Stevens, B., and Bao, J.: Variations of tropical lapse rates in climate models and their implications for upper-tropospheric warming, *J. Climate*, 34, 9747–9761, <https://doi.org/10.1175/JCLI-D-21-0196.1>, 2021.
- Kim, D.-W., Zapponini, M., Jung, T., Kim, M.-H., Koldunov, N., Puthiyaveetil, N., Sharma, S., Sidorenko, D., Sinha, N., Streffing, J., Timmermann, A., Semmler, T., and Park, W.: Accelerating Southern Ocean Sea Ice Decline due to Cloud-Climate Feedbacks, *Research Square* [preprint], <https://doi.org/10.21203/rs.3.rs-7325013/v1>, 2026.
- Koldunov, N. V., Aizinger, V., Rakowsky, N., Scholz, P., Sidorenko, D., Danilov, S., and Jung, T.: Scalability and some optimization of the Finite-volume Sea ice–Ocean Model, Version 2.0 (FESOM2), *Geosci. Model Dev.*, 12, 3991–4012, <https://doi.org/10.5194/gmd-12-3991-2019>, 2019a.
- Koldunov, N., Danilov, S., Sidorenko, D., Hutter, N., Losch, M., Goessling, H., Rakowsky, N., Scholz, P., Sein, D., Wang, Q., and Jung, T.: Fast EVP solutions in a high-resolution sea ice model, *J. Adv. Model. Earth Sy.*, 11, 1269–1284, <https://doi.org/10.1029/2018MS001485>, 2019b.
- Komen, G. J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., and Janssen, P.: *Dynamics and modelling of ocean waves*, Cambridge University Press, UK, <https://doi.org/10.1017/CBO9780511628955>, 1996.
- Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, *Rev. Geophys.*, 32, 363–403, <https://doi.org/10.1029/94RG01872>, 1994.
- Lauer, A., Bock, L., Hassler, B., Schröder, M., and Stengel, M.: Cloud climatologies from global climate models – A comparison of CMIP5 and CMIP6 models with satellite data, *J. Climate*, 36, 281–311, <https://doi.org/10.1175/JCLI-D-22-0181.1>, 2023.
- Lavergne, T. and Down, E.: A climate data record of year-round global sea-ice drift from the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF), *Earth Syst. Sci. Data*, 15, 5807–5834, <https://doi.org/10.5194/essd-15-5807-2023>, 2023.
- Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., Bricaud, C., Hamon, M., Le Galoudec, O., Regnier, C., Candela, T., Testut, C.-E., Gasparin, F., Ruggiero, G., Benkiran, M., Drillet, Y., and Le Traon, P.-Y.: The Copernicus global 1/12 oceanic and sea ice GLORYS12 reanalysis, *Front. Earth Sci.*, 9, 698876, <https://doi.org/10.3389/feart.2021.698876>, 2021.
- Li, D., Chang, P., Yeager, S. G., Danabasoglu, G., Castruccio, F. S., Small, J., Wang, H., Zhang, Q., and Gopal, A.: The impact of horizontal resolution on projected sea-level rise along US east continental shelf with the community earth system model, *J. Adv. Model. Earth Sy.*, 14, e2021MS002868, <https://doi.org/10.1029/2021MS002868>, 2022.
- Li, X., Wang, Q., Danilov, S., Koldunov, N., Liu, C., Müller, V., Sidorenko, D., and Jung, T.: Eddy activity in the Arctic Ocean projected to surge in a warming world, *Nat. Clim. Change*, 14, 156–162, 2024.
- Marshall, J. and Speer, K.: Closure of the meridional overturning circulation through Southern Ocean upwelling, *Nat. Geosci.*, 5, 171–180, <https://doi.org/10.1038/ngeo1391>, 2012.
- Mercator Ocean International: Global Ocean Physics Reanalysis, E.U. Copernicus Marine Service Information (CMEMS) Marine Data Store (MDS) [data set], <https://doi.org/10.48670/moi-00021>, 2018.
- Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., Fraser, P. J., Montzka, S. A., Rayner, P. J., Trudinger, C. M., Krummel, P. B., Beyerle, U., Canadell, J. G., Daniel, J. S., Enting, I. G., Law, R. M., Lunder, C. R., O'Doherty, S., Prinn, R. G., Reimann, S., Rubino, M., Velders, G. J. M., Vollmer, M. K., Wang, R. H. J., and Weiss, R.: Historical greenhouse gas concentrations for climate modelling (CMIP6), *Geosci. Model Dev.*, 10, 2057–2116, <https://doi.org/10.5194/gmd-10-2057-2017>, 2017.
- Meredith, M. P., Woodworth, P. L., Chereskin, T. K., Marshall, D. P., Allison, L. C., Bigg, G. R., Donohue, K., Heywood, K. J., Hughes, C. W., Hibbert, A., McC. Hogg, A., Johnson, H. J., Jullion, L., King, B. A., Leach, H., Lenn, Y. D., Morales Maqueda, M. A., Munday, D. A., Naveira Garabato, A. C., Provost, C., Sallée, J.-B., and Sprintall, J.: Sustained monitoring of the Southern Ocean at Drake Passage: Past achievements and future priorities, *Rev. Geophys.*, 49, <https://doi.org/10.1029/2010RG000348>, 2011.
- Mitchell, D. M., Lo, E. Y. T., Seviour, W. J. M., Haimberger, L., and Polvani, L. M.: The vertical profile of recent tropical temperature trends: Persistent model biases in the context of internal variability, *Environ. Res. Lett.*, 15, 1040b4, <https://doi.org/10.1088/1748-9326/ab9af7>, 2020.
- Mmame, B. and Ngongondo, C.: Evaluation of CMIP6 model skills in simulating tropical climate extremes over Malawi, Southern Africa, *Modeling Earth Systems and Environment*, 10, 1695–1709, <https://doi.org/10.1007/s40808-023-01867-3>, 2024.
- Moat, B. I., Smeed, D. A., Rayner, D., Johns, W. E., Smith, R., Volkov, D., Elipot, S., Petit, T., Kajtar, J., Baringer, M. O., and Collins, J.: Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2023 (v2023.1), British Oceanographic Data Centre – Natural Environment Research Council [data set], <https://doi.org/10.5285/223b34a3-2dc5-c945-e063-7086abc0f274>, 2024.
- Mohrmann, M., Heuzé, C., and Swart, S.: Southern Ocean polynyas in CMIP6 models, *The Cryosphere*, 15, 4281–4313, <https://doi.org/10.5194/tc-15-4281-2021>, 2021.
- Monterey, G. I. and Levitus, S.: Seasonal variability of mixed layer depth for the world ocean, NOAA (NESDIS se-

- ries), https://repository.library.noaa.gov/view/noaa/49153/noaa_49153_DS1.pdf (last access: 3 April 2025), 1997.
- Moon, J.-Y., Streffing, J., Lee, S.-S., Semmler, T., Andrés-Martínez, M., Chen, J., Cho, E.-B., Chu, J.-E., Franzke, C. L. E., Gärtner, J. P., Ghosh, R., Hegewald, J., Hong, S., Kim, D.-W., Koldunov, N., Lee, J.-Y., Lin, Z., Liu, C., Loza, S. N., Park, W., Roh, W., Sein, D. V., Sharma, S., Sidorenko, D., Son, J.-H., Stuecker, M. F., Wang, Q., Yi, G., Zappaonini, M., Jung, T., and Timmermann, A.: Earth's future climate and its variability simulated at 9 km global resolution, *Earth Syst. Dynam.*, 16, 1103–1134, <https://doi.org/10.5194/esd-16-1103-2025>, 2025.
- Moreno-Chamarro, E., Caron, L.-P., Loosveldt Tomas, S., Vegas-Regidor, J., Gutjahr, O., Moine, M.-P., Putrasahan, D., Roberts, C. D., Roberts, M. J., Senan, R., Terray, L., Tourigny, E., and Vidale, P. L.: Impact of increased resolution on long-standing biases in HighResMIP-PRIMAVERA climate models, *Geosci. Model Dev.*, 15, 269–289, <https://doi.org/10.5194/gmd-15-269-2022>, 2022.
- Muilwijk, M., Hattermann, T., Martin, T., and Granskog, M. A.: Future sea ice weakening amplifies wind-driven trends in surface stress and Arctic Ocean spin-up, *Nat. Commun.*, 15, 6889, <https://doi.org/10.1038/s41467-024-50874-0>, 2024.
- Nakayama, Y., Malyarenko, A., Zhang, H., Wang, O., Auger, M., Nie, Y., Fenty, I., Mazloff, M., Köhl, A., and Menemenlis, D.: Evaluation of MITgcm-based ocean reanalyses for the Southern Ocean, *Geosci. Model Dev.*, 17, 8613–8638, <https://doi.org/10.5194/gmd-17-8613-2024>, 2024.
- NASA/LARC/SD/ASDC: CERES Energy Balanced and Filled (EBAF) TOA and surface monthly means data in netCDF edition 4.1. NASA, https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF_L3B.004.1, 2019.
- Notz, D. and SIMIP Community: Arctic sea ice in CMIP6, *Geophys. Res. Lett.*, 47, e2019GL086749, <https://doi.org/10.1029/2019GL086749>, 2020.
- Pimonsree, S., Kamworapan, S., Gheewala, S. H., Thongbhakdi, A., and Prueksakorn, K.: Evaluation of CMIP6 GCMs performance to simulate precipitation over Southeast Asia, *Atmos. Res.*, 282, 106522, <https://doi.org/10.1016/j.atmosres.2022.106522>, 2023.
- Platnick, S., Ackerman, S., King, M., Meyer, K., Menzel, W. P., Holz, R. E., Baum, B. A., and Yang, P.: MODIS atmosphere L2 cloud product (06_L2), NASA MODIS Adaptive Processing System, Goddard Space Flight Center [data set], https://doi.org/10.5067/MODIS/MOD06_L2.006, 2015.
- Putrasahan, D. A., Lohmann, K., von Storch, J.-S., Jungclaus, J. H., Gutjahr, O., and Haak, H.: Surface flux drivers for the slowdown of the Atlantic meridional overturning circulation in a high-resolution global coupled climate model, *J. Adv. Model. Earth Sy.*, 11, 1349–1363, <https://doi.org/10.1029/2018MS001447>, 2019.
- Rabazanahary, T., Mirindra, F., Zhai, J., and Andrianarimanana, M. H.: Performance evaluation of CMIP6 in simulating extreme precipitation in Madagascar, *Theor. Appl. Climatol.*, 155, 4089–4100, <https://doi.org/10.1007/s00704-024-04868-z>, 2024.
- Rackow, T., Sein, D. V., Semmler, T., Danilov, S., Koldunov, N. V., Sidorenko, D., Wang, Q., and Jung, T.: Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations with AWI-CM1.0, *Geosci. Model Dev.*, 12, 2635–2656, <https://doi.org/10.5194/gmd-12-2635-2019>, 2019.
- Rackow, T., Danilov, S., Goessling, H. F., Hellmer, H. H., Sein, D. V., Semmler, T., Sidorenko, D., and Jung, T.: Delayed Antarctic sea-ice decline in high-resolution climate change simulations, *Nat. Commun.*, 13, 637, <https://doi.org/10.1038/s41467-022-28259-y>, 2022.
- Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., Bechtold, P., Beyer, S., Bidlot, J., Boussetta, S., Deconinck, W., Diamantakis, M., Dueben, P., Dutra, E., Forbes, R., Ghosh, R., Goessling, H. F., Hadade, I., Hegewald, J., Jung, T., Keeley, S., Kluff, L., Koldunov, N., Koldunov, A., Kölling, T., Kousal, J., Kühnlein, C., Maciel, P., Mogensen, K., Quintino, T., Polichtchouk, I., Reuter, B., Sármany, D., Scholz, P., Sidorenko, D., Streffing, J., Sützl, B., Takasuka, D., Tietsche, S., Valentini, M., Vannière, B., Wedi, N., Zampieri, L., and Ziemann, F.: Multi-year simulations at kilometre scale with the Integrated Forecasting System coupled to FESOM2.5 and NEMOv3.4, *Geosci. Model Dev.*, 18, 33–69, <https://doi.org/10.5194/gmd-18-33-2025>, 2025.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.-Atmos.*, 108, <https://doi.org/10.1029/2002JD002670>, 2003.
- Redi, M. H.: Oceanic isopycnal mixing by coordinate rotation, *J. Phys. Oceanogr.*, 12, 1154–1158, [https://doi.org/10.1175/1520-0485\(1982\)012<1154:OIMBCR>2.0.CO;2](https://doi.org/10.1175/1520-0485(1982)012<1154:OIMBCR>2.0.CO;2), 1982.
- Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, *B. Am. Meteorol. Soc.*, 89, 303–312, 2008.
- Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., and Simpson, I. R.: Progress in simulating the quasi-biennial oscillation in CMIP models, *J. Geophys. Res.-Atmos.*, 125, e2019JD032362, <https://doi.org/10.1029/2019JD032362>, 2020.
- Ridderinkhof, H., Van der Werf, P. M., Ullgren, J. E., Van Aken, H. M., Van Leeuwen, P. J., and De Ruijter, W. P. M.: Seasonal and interannual variability in the Mozambique Channel from moored current observations, *J. Geophys. Res.-Oceans*, 115, <https://doi.org/10.1029/2009JC005619>, 2010.
- Roach, L. A., Dörr, J., Holmes, C. R., Massonnet, F., Blockley, E. W., Notz, D., Rackow, T., Raphael, M. N., O'Farrell, S. P., Bailey, D. A., and Bitz, C. M.: Antarctic sea ice area in CMIP6, *Geophys. Res. Lett.*, 47, e2019GL086729, <https://doi.org/10.1029/2019GL086729>, 2020.
- Roberts, M. J., Reed, K. A., Bao, Q., Barsugli, J. J., Camargo, S. J., Caron, L.-P., Chang, P., Chen, C.-T., Christensen, H. M., Danabasoglu, G., Frenger, I., Fučkar, N. S., ul Hasson, S., Hewitt, H. T., Huang, H., Kim, D., Kodama, C., Lai, M., Leung, L.-Y. R., Mizuta, R., Nobre, P., Ortega, P., Paquin, D., Roberts, C. D., Scoccimarro, E., Seddon, J., Treguier, A. M., Tu, C.-Y., Ullrich, P. A., Vidale, P. L., Wehner, M. F., Zarzycki, C. M., Zhang, B., Zhang, W., and Zhao, M.: High-Resolution Model Intercomparison Project phase 2 (HighResMIP2) towards CMIP7, *Geosci. Model Dev.*, 18, 1307–1332, <https://doi.org/10.5194/gmd-18-1307-2025>, 2025.
- Scholz, P., Sidorenko, D., Danilov, S., Wang, Q., Koldunov, N., Sein, D., and Jung, T.: Assessment of the Finite-VolumE Sea Ice–Ocean Model (FESOM2.0) – Part 2: Partial bottom cells, embedded sea ice and vertical mixing library CVMix, *Geosci.*

- Model Dev., 15, 335–363, <https://doi.org/10.5194/gmd-15-335-2022>, 2022.
- Scholz, P., Sidorenko, D., Streffing, J., Hegewald, J., Koldunov, N., Cheedela, S. K., Rackow, T., Loch, W. J., Gierz, P., Ackermann, L., Deconinck, W., Butzin, M., Song, P., Gürses, O., Heukamp, F., Degregori, E., Hinck, S., Beyer, S., Andrés-Martínez, M., Wilms, A., Vitali, E., Kinoshita, B. P., Goessling, H. F., Zappaonini, M., Müller, V., Koldunov, A., Guibert, D., Wang, Q., and Okzan, U.: mzappaonini/fesom2: FESOM2 version used for production runs in Zappaonini et al. (2025), Zenodo [code], <https://doi.org/10.5281/zenodo.17961041>, 2025.
- Sein, D., Danilov, S., Biastoch, A., Durgadoo, J. V., Sidorenko, D., Harig, S., and Wang, Q.: Designing variable ocean model resolution based on the observed ocean variability, *J. Adv. Model. Earth Sy.*, 8, 904–916, <https://doi.org/10.1002/2016MS000650>, 2016.
- Sein, D., Koldunov, N., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., Rackow, T., Cabos, W., and Jung, T.: Ocean modeling on a mesh with resolution following the local Rossby radius, *J. Adv. Model. Earth Sy.*, 9, 2601–2614, <https://doi.org/10.1002/2017MS001099>, 2017.
- Shi, Q., Yang, Q., Mu, L., Wang, J., Massonnet, F., and Mazloff, M. R.: Evaluation of sea-ice thickness from four reanalyses in the Antarctic Weddell Sea, *The Cryosphere*, 15, 31–47, <https://doi.org/10.5194/tc-15-31-2021>, 2021.
- Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., Stevens, B., Stocker, T., and Teutsch, G.: Ambitious partnership needed for reliable climate prediction, *Nat. Clim. Change*, 12, 499–503, <https://doi.org/10.1038/s41558-022-01384-8>, 2022.
- Song, J.-N., Fu, G., Xu, Y., Han, Z.-Y., Sun, Q.-Z., and Wang, H.: Assessment of the capability of CMIP6 global climate models to simulate Arctic cyclones, *Advances in Climate Change Research*, 12, 660–676, <https://doi.org/10.1016/j.accre.2021.07.007>, 2021.
- Soto-Navarro, J., Criado-Aldeanueva, F., García-Lafuente, J., and Sánchez-Román, A.: Estimation of the Atlantic inflow through the Strait of Gibraltar from climatological and in situ data, *J. Geophys. Res.-Oceans*, 115, <https://doi.org/10.1029/2010JC006302>, 2010.
- Steele, M., Morley, R., and Ermold, W.: PHC: A global ocean hydrography with a high-quality Arctic Ocean, *J. Climate*, 14, 2079–2087, [https://doi.org/10.1175/1520-0442\(2001\)014<2079:PAGOHW>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<2079:PAGOHW>2.0.CO;2), 2001.
- Streffing, J., Zappaonini, M., Andrés-Martínez, M., and Winters, N.: JanStreffing/cmpitool: v1.2 (v1.2), Zenodo [code], <https://doi.org/10.5281/zenodo.16961790>, 2025.
- Streffing, J., Sidorenko, D., Semmler, T., Zampieri, L., Scholz, P., Andrés-Martínez, M., Koldunov, N., Rackow, T., Kjellsson, J., Goessling, H., Athanase, M., Wang, Q., Hegewald, J., Sein, D. V., Mu, L., Fladrich, U., Barbi, D., Gierz, P., Danilov, S., Juricke, S., Lohmann, G., and Jung, T.: AWI-CM3 coupled climate model: description and evaluation experiments for a prototype post-CMIP6 model, *Geosci. Model Dev.*, 15, 6399–6427, <https://doi.org/10.5194/gmd-15-6399-2022>, 2022.
- Taylor, G. P., Loikith, P. C., Aragon, C. M., Lee, H., and Waliser, D. E.: CMIP6 model fidelity at simulating large-scale atmospheric circulation patterns and associated temperature and precipitation over the Pacific Northwest, *Clim. Dynam.*, 60, 2199–2218, <https://doi.org/10.1007/s00382-022-06410-1>, 2023.
- Tian, B. and Dong, X.: The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation, *Geophys. Res. Lett.*, 47, e2020GL087232, <https://doi.org/10.1029/2020GL087232>, 2020.
- Tian, T., Yang, S., Høyer, J., Nielsen-Englyst, P., and Singha, S.: Cooler Arctic surface temperatures simulated by climate models are closer to satellite-based data than the ERA5 reanalysis, *Communications Earth & Environment*, 5, 111, <https://doi.org/10.1038/s43247-024-01276-z>, 2024.
- Timmermann, R. and Beckmann, A.: Parameterization of vertical mixing in the Weddell Sea, *Ocean Model.*, 6, 83–100, [https://doi.org/10.1016/S1463-5003\(02\)00061-6](https://doi.org/10.1016/S1463-5003(02)00061-6), 2004.
- Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, *Science Advances*, 6, eaaz9549, <https://doi.org/10.1126/sciadv.aaz9549>, 2020.
- von Storch, J.-S., Haak, H., Hertwig, E., and Fast, I.: Vertical heat and salt fluxes due to resolved and parameterized meso-scale eddies, *Ocean Model.*, 108, 1–19, <https://doi.org/10.1016/j.ocemod.2016.10.001>, 2016.
- Walden, V. P., Hudson, S. R., Cohen, L., Murphy, S. Y., and Granskog, M. A.: Atmospheric components of the surface energy budget over young sea ice: Results from the N-ICE2015 campaign, *J. Geophys. Res.-Atmos.*, 122, 8427–8446, <https://doi.org/10.1002/2016JD026091>, 2017.
- Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee III, R. B., Smith, G. L., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment, *B. Am. Meteorol. Soc.*, 77, 853–868, [https://doi.org/10.1175/1520-0477\(1996\)077<0853:CATERE>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2), 1996.
- Wild, M.: The global energy balance as represented in CMIP6 climate models, *Clim. Dynam.*, 55, 553–577, <https://doi.org/10.1007/s00382-020-05282-7>, 2020.
- Woodgate, R. A.: Increases in the Pacific inflow to the Arctic from 1990 to 2015, and insights into seasonal trends and driving mechanisms from year-round Bering Strait mooring data, *Prog. Oceanogr.*, 160, 124–154, <https://doi.org/10.1016/j.pocean.2017.12.007>, 2018.
- Woodgate, R. A., Aagaard, K., and Weingartner, T. J.: Monthly temperature, salinity, and transport variability of the Bering Strait through flow, *Geophys. Res. Lett.*, 32, <https://doi.org/10.1029/2004GL021880>, 2005.
- Yepes-Arbós, X., van den Oord, G., Acosta, M. C., and Carver, G. D.: Evaluation and optimisation of the I/O scalability for the next generation of Earth system models: IFS CY43R3 and XIOS 2.0 integration as a case study, *Geosci. Model Dev.*, 15, 379–394, <https://doi.org/10.5194/gmd-15-379-2022>, 2022.
- Zappaonini, M.: AWI-CM3.1 historical and ssp585 scenario in TCo95-CORE2 setup, Zenodo [data set], <https://doi.org/10.5281/zenodo.17377570>, 2025a.
- Zappaonini, M.: AWI-CM3.1 historical and ssp585 scenario in TCo319-DART setup, Zenodo [data set], <https://doi.org/10.5281/zenodo.17397357>, 2025b.
- Zappaonini, M.: mzappaonini/AWICM3_ResolutionSensitivity_paper: Notebooks_AWI-CM3_resolutions, Zenodo [code], <https://doi.org/10.5281/zenodo.17996211>, 2025c.
- Zappaonini, M. and Goessling, H.: Atmospheric destabilization leads to Arctic Ocean winter surface wind intensi-

- fication, *Communications Earth & Environment*, 5, 262, <https://doi.org/10.1038/s43247-024-01428-1>, 2024.
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, Sand Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of higher climate sensitivity in CMIP6 models, *Geophys. Res. Lett.*, 47, e2019GL085782, <https://doi.org/10.1029/2019GL085782>, 2020.
- Zhang, B., Song, S., Wang, Hand Guo, T., and Ding, Y.: Evaluation of the performance of CMIP6 models in simulating extreme precipitation and its projected changes in global climate regions, *Nat. Hazards*, 121, 1737–1763, <https://doi.org/10.1007/s11069-024-06850-4>, 2025a.
- Zhang, R., Sun, S., Chen, Z., and Wu, L.: Limited influence of the Agulhas leakage on the Atlantic Meridional Overturning Circulation under present-day climate conditions, *Communications Earth & Environment*, 6, 133, <https://doi.org/10.1038/s43247-025-02097-4>, 2025b.
- Zhao, L., Wang, Y., Zhao, C., Dong, X., and Yung, Y. L.: Compensating errors in cloud radiative and physical properties over the Southern Ocean in the CMIP6 climate models, *Adv. Atmos. Sci.*, 39, 2156–2171, <https://doi.org/10.1007/s00376-022-2036-z>, 2022.