Author's response to reviews of "Assessing resolution sensitivity in coupled climate simulations with AWI-CM3"

We thank the reviewers for their comments on our manuscript. We are glad to see that our manuscript has been found of interest. The point-by-point replies to the comments are below, with reviewer comments in black and our responses in blue.

Reviewer #1

General comments:

The paper provides a useful overview of the coupled AWI-CM3 model, providing a near direct comparison between two model resolutions, as well as evaluation against observations/re-analysis and CMIP6. Particular attention is given to polar climate, where increased resolution has a larger impact on mesoscale processes unresolved by coarser resolution. The comparison is between a pair of simulations from 1950 - 2100, with most of the emphasis on evaluation of biases relative to present-day observations or reanalysis, rather than on future projections. Overall this is a well-written and clearly understandable manuscript that provides a very useful benchmark for the AWI-CM3 model and the impact of increased resolution.

We thank the reviewer for this positive and constructive comment.

Specific comments:

Each model is run for 250 years from initialization in repeat 1950 conditions, with the first 100 years considered spinup and subsequent 150 years as control. It would be useful to have some mention of the model's level of equilibrium during this control period, and/or see metrics. Figure 3 (now Figure 4) shows some of this; the global SST and 2m air temperatures seem to be in equilibrium in the control, but maybe some of the sea ice metrics have a trend during the control? Is the ocean steadily gaining or losing heat during the control? In Figure 7d (now Figure 8d) the AMOC time series appear to be declining from 1950, was this trend also in the control?

We have added a new subsection at the beginning of the Results ("Model drift and equilibration") to explicitly assess the degree of equilibration of both model configurations during the control period. This includes an analysis of the TOA and surface energy imbalances, as well as the vertical structure of the global-mean ocean temperature bias. The revised analysis shows that the LR configuration reaches a quasi-equilibrated state by the beginning of the control period, with a small and stable positive TOA imbalance, whereas the MR simulation exhibits a residual imbalance indicative of a weak but ongoing ocean adjustment. The LR simulation is also characterized by more pronounced trends in ocean temperature, particularly during the spin-up phase, compared to the higher-resolution configuration. However, when focusing on the period from 1950 onwards, both simulations exhibit only weak drift in global-mean SST and near-surface air temperature, as well as relatively stable sea ice extent during the control period (included in Figure 4). The ocean heat content evolution is consistent with the diagnosed radiative imbalances, indicating a slight heat uptake in LR and a weak heat loss in MR. Regarding the AMOC, we have extended the analysis to include the control simulation (see updated Figure 8). This shows that the declining trend in the northward branch is primarily associated with the transient forcing and is not present in the control simulation.

The additional paragraph and the revised figures are reported here in the following.

2.1 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 (Figure 2a,b). After spin-up, the imbalance stabilizes at approximately +0.2 W m^{-2} in the low-resolution configuration, 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 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 (Figure 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 (Figure 4).

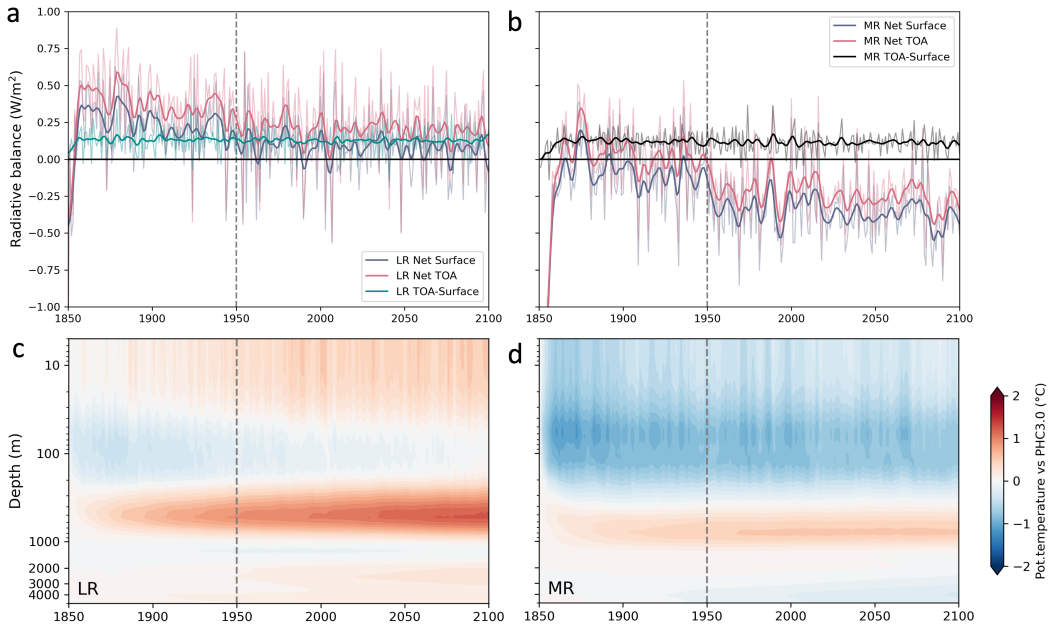

Figure 2. (a,b) Net radiative imbalance at the top of the atmosphere (TOA) and at the surface, together with their difference, representing the atmospheric column energy balance, during the spin-up and control simulations. Positive (negative) values indicate downward (upward) fluxes. (c,d) Hovmöller 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.

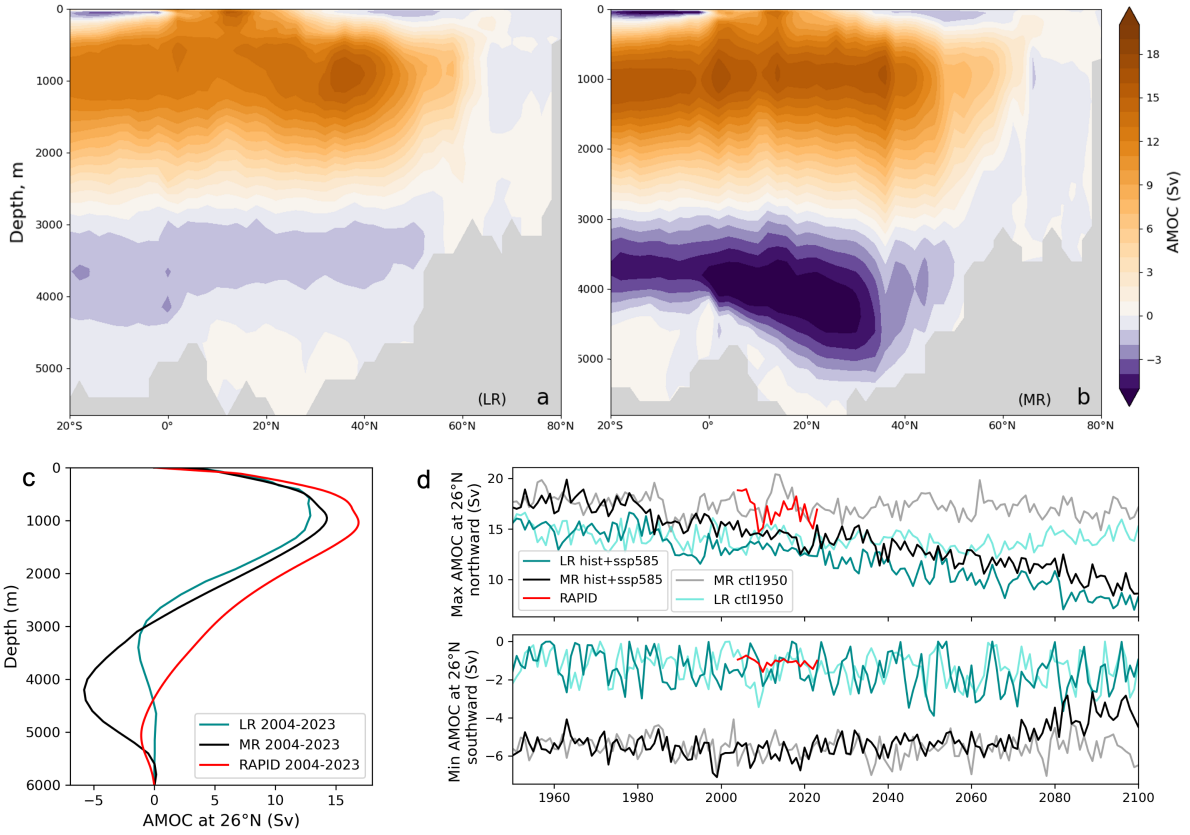

Figure 8. Atlantic Meridional Overturning Circulation (AMOC) overturning streamfunction averaged over the period 1979–2014 for (a) the low-resolution simulation (LR) and (b) the medium-resolution simulation (MR). (c) Comparison of AMOC overturning streamfunctions from AWI-CM3 simulations and RAPID-MOC array (**?**) 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 Sv = 106 m^3 s^{-1})

Reviewer #2

This paper investigates the impact of model resolution in coupled simulations using the AWI-CM3 model, which is one of the CMIP6-participating models. The model was run in two configurations: a lower-resolution setup (LR; atmosphere: 102 to 120 km with 91 vertical levels; ocean: CORE2, 20 to 100 km) and a higher, medium-resolution setup (MR; atmosphere: 32 to 38 km with 137 levels; ocean: DART, 4.5 to 25 km). The higher resolution permits ocean eddy-resolving capability. The paper covers a variety of evaluations to examine performance changes due to the increased resolution. The authors find that increased resolution leads to improved model performance overall. I believe the paper is well written and the scope fits the journal. However, before publication, readers would benefit from a revised manuscript, including the following.

We thank the reviewer for the careful and constructive evaluation of our manuscript, and for the positive assessment of its scope and clarity. We appreciate the helpful suggestions provided. Below, we address each comment in detail.

Section 2: What is the simulation period? I assume that is identical to the CMIP6 Historical experiment protocol, but it may be worth noting it clearly in the text.

We thank the reviewer for this suggestion. The simulation protocol was already described in the manuscript (lines 108–114); however, we have now introduced a dedicated subsection to improve the accessibility of the information.

Table 1: I suggest adding the equivalent resolution information in km for the atmospheric and ocean grids, which will help improving the clarity of the configuration.

The horizontal grid spacing ranges (in km) have now been included in the table.

Fig. 1: I suggest use “LR” and “MR” or “lower” and “higher” for the 2nd and 3rd column titles, as the readers may not familiar with “CORE2” and “DART” terminologies thus have to revisit the text where they were defined.

We have revised the figure and its caption to improve clarity.

Fig. 2 (Now Fig.3): I suggest adding a color bar. Also, readers would benefit from knowing exact variables used in the figure. Some abbreviated variable names on the y-axis may not be very straightforward especially for those who are not in the field (e.g., mlotst, thetao, so, etc), although some were described in the text (line 155-170). I suggest providing full names of the variables (recommended) in a comprehensive way (like a table in the supplementary) and point to where they could be found from the figure caption.

We thank the reviewer for the suggestions to improve the readability of the heatmaps. We have now added a colorbar to the figure and included a table mapping variable names in the Supplementary Figures.

Line 151-153: Regarding the overall score, some arbitrary is unavoidable depending the selection of variables and regions, and there has been long debates about defining the “overall score” in the community. I recommend discussing about such limitation

of the overall score.

We agree with the reviewer that the overall score should be interpreted with caution. We have therefore added a clarification in the manuscript (lines 209-213): “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.”

Fig 6 (Now Fig.7): I find a mismatch between the figure and the interpretating text, and it puzzles me. It looks like the color bar to be the opposite of the interpretation of the result in the section 3.3. For example, lines 254-255 describes about the observed sea ice loss in the central Arctic (Fig. 6g) where in the figure, most of the Arctic sea is colored in red, in which according to the color bar, it indicates positive SIC trend, thus increasing sea ice. I encourage the authors double check the figure to make sure what it shows.

We thank the reviewer for pointing out this issue. The colorbar was inadvertently reversed and has now been corrected.

Fig 11 (Now Fig.12). I recommend using the same configuration name (LR, MR) throughout the paper for consistency. I also wonder the lines from LR, MR, and OBS could be superimposed to same space and make comparison easier.

We thank the reviewer for this suggestion. We have updated the configuration names and colors to ensure consistency with the rest of the analysis. We have, however, retained the three-column layout. As the time series can be noisy, overlaying them can make the comparison less clear. We believe that the current layout clearly highlights the main point of the comparison: the correspondence in variability in the MR configuration, which is also present in reanalysis and observations, but absent in the LR configuration.

The revised figures and the additional table are reported here in the following.

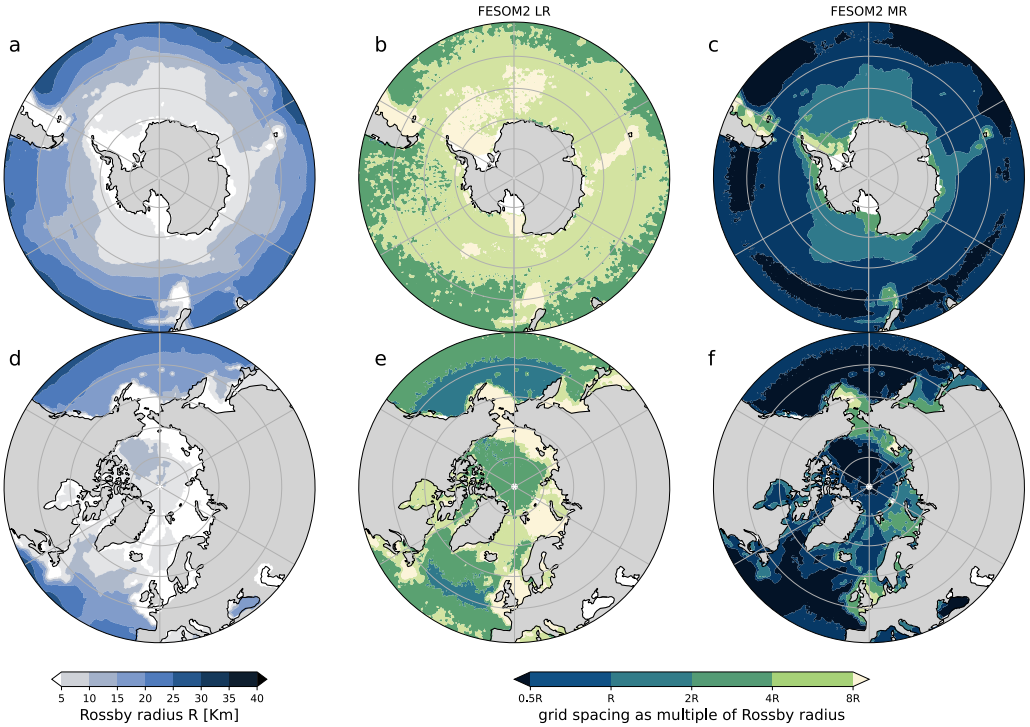

Figure 1. Local Rossby radius of deformation (a,d) estimated as $R = (\pi f)^{-1} \int_{-H}^{0} N(z)dz$ where f is the Coriolis parameter, H is the ocean floor, and N is the Brunt-Väisälä frequency (as in Sein et al., 2016). Ocean horizontal grid spacing as a multiple of the local Rossby radius for the (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 (102–120km)	CORE2L47 (20–100km)	XIOS parallel	1664	73.80	541
MR	TCo319L137 (32–38km)	DARTL80 (4.5–25km)	XIOS parallel	8576	3.89	52 911

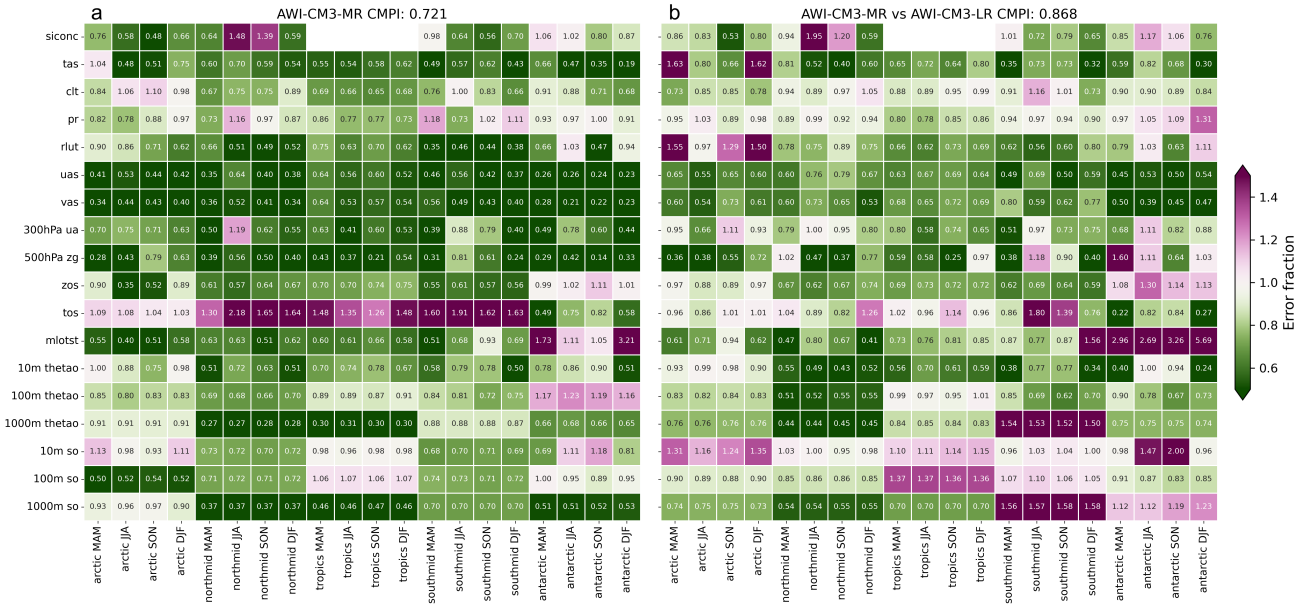

Figure 3. (a) Performance indices showing the absolute error of the AWI-CM3-MR historic climatology (1989–2014) expressed as a fraction of the CMIP6 multi-model mean absolute error (30 models; (Streffing, 2022; Reichler and Kim, 2008)). Values below (above) 1 indicate smaller (larger) biases compared to the CMIP6 average. Panel (a) compares AWI-CM3-MR against CMIP6, while Figure 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

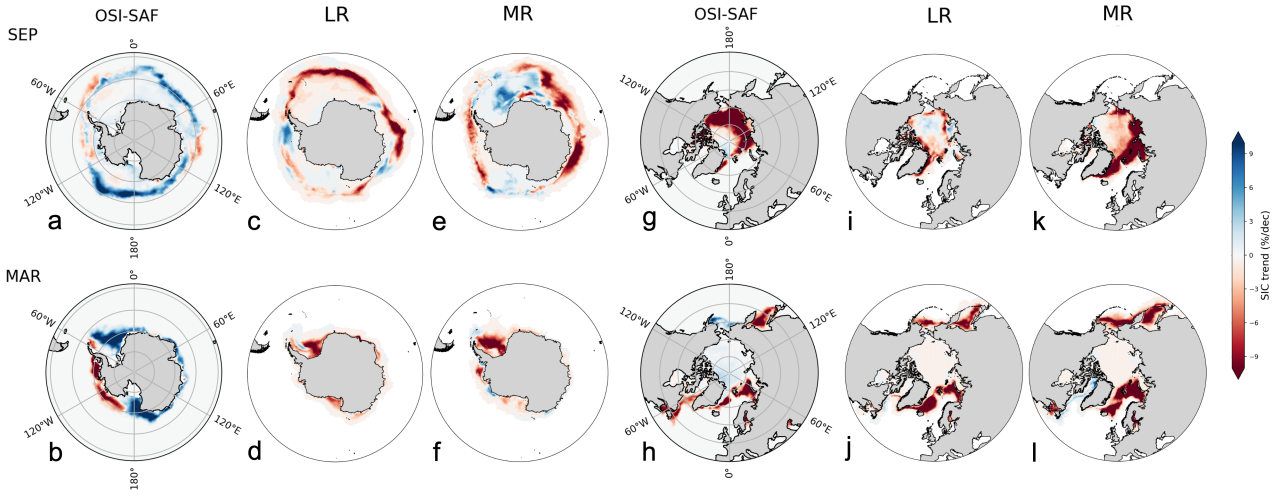

Figure 7. Trends in March and September sea ice concentration from 1979 to 2014 are shown for both hemispheres, based on OSI-SAF observations (a,b,g,h) and simulations from the low-resolution (LR, c,d,i,j) and medium-resolution (MR, e,f,k,l) model configurations. Although the two observational datasets display some differences in the time series shown in Figure 3b–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.

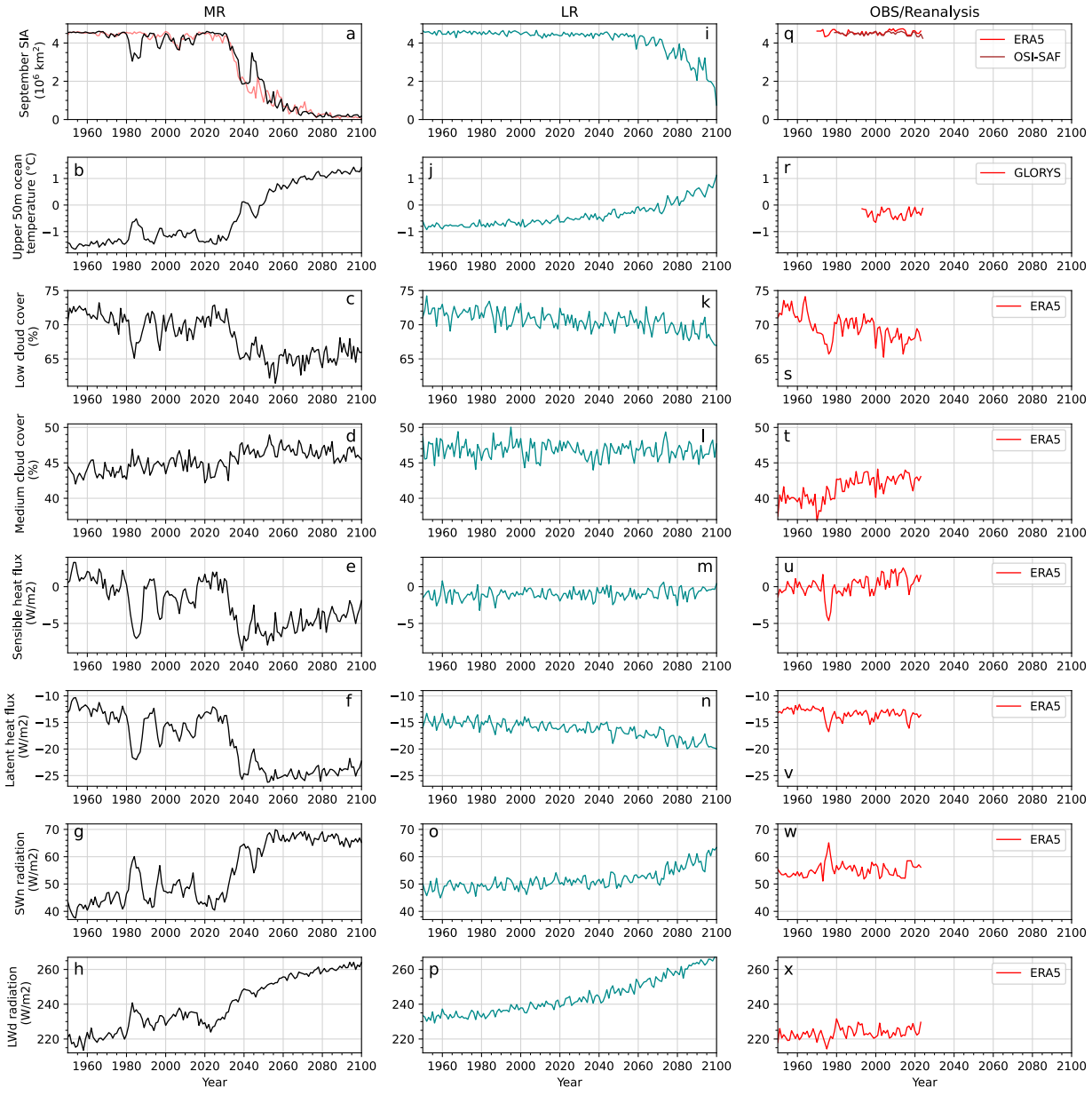

Figure 12. Time series of key variables involved in atmosphere–ocean interaction feedbacks. Panels (q–x) show September mean sea ice area and annual mean of upper 50 m ocean temperature, low- and mid-level cloud cover, surface sensible and latent heat fluxes, surface net shortwave radiation (SWn) and surface downward longwave radiation (LWd) over the Weddell Sea region (50°W–30°E, 60°S–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).

Appendix A: Supplementary Figures

CMIP standard name	Extended variable name (unit)
siconc	Sea ice concentration (%)
tas	Near-surface air temperature (K)
clt	Total cloud cover (%)
pr	Precipitation flux (kg m^{-2} 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})
300hPa ua	Zonal wind at 300 hPa (m s^{-1})
500hPa zg	Geopotential height at 500 hPa (m)
zos	Sea surface height above geoid (m)
tos	Sea surface temperature (°C)
mlotst	Ocean mixed layer thickness (m)
10m thetao	Ocean potential temperature at 10 m depth (°C)
100m thetao	Ocean potential temperature at 100 m depth (°C)
1000m thetao	Ocean potential temperature at 1000 m depth (°C)
10m so	Ocean salinity at 10 m depth (1e-3)
100m so	Ocean salinity at 100 m depth (1e-3)
1000m so	Ocean salinity at 1000 m depth (1e-3)

Table A1. List of CMIP variables used in Figure 3 and Figure **??** given together with their corresponding standard names and units.