

RESEARCH ARTICLE

A multimodel intercomparison study of variable-resolution global models with grid refinement over the Arctic and Antarctic

Lise Seland Graff¹  | Dörthe Handorf² | Raphael Köhler² | Xavier Levine³ | René R. Wijngaard⁴ | Gareth J. Marshall⁵ | Alexey Yu. Karpechko⁶ | Ryan S. Williams⁵ | Priscilla A. Mooney³

¹Norwegian Meteorological Institute, Oslo, Norway

²Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

³NORCE Research AS, Bjerknes Centre for Climate Research, Bergen, Norway

⁴Utrecht University, Utrecht, Netherlands

⁵British Antarctic Survey, Cambridge, UK

⁶Finnish Meteorological Institute, Helsinki, Finland

Correspondence

Lise Seland Graff, Development Center for Weather Forecasting, Norwegian Meteorological Institute, P.O. Box 43 Blindern, N-0313 Oslo, Norway.
Email: lise.s.graff@met.no

Funding information

European Commission H2020 project PolarRES, Grant/Award Number: 101003590; Research Council of Norway project BASIC, Grant/Award Number: 325440

Abstract

The polar regions are undergoing profound changes, but climate models still struggle to accurately represent key polar climate processes. Exploring how increased horizontal resolution affects model biases in the polar regions and lower latitudes, we performed a coordinated set of global model simulations for the recent past with grid refinement in the polar regions, with the finest resolution implemented poleward of 60°. We employed three state-of-the-art models: the Community Earth System Model, the Icosahedral Nonhydrostatic model, and the Model for Prediction Across Scales. The result is the first coordinated multimodel ensemble with variable-resolution models with refinement over the polar regions. This work documents the common experiment protocol and presents the first results, focusing on temperature, zonal wind, and turbulent surface fluxes. The experiments were configured with fully coupled atmosphere and land components and prescribed sea-surface temperatures and sea-ice concentrations. Forcings for greenhouse gases and aerosols were based on the historical experiment from the sixth phase of the Coupled Model Intercomparison Project. We find that the influence of the refinement, assessed by comparing simulations with refined and uniform resolution, can vary considerably across models, hemispheres, and seasons. The most consistent improvements occur in wintertime upper-tropospheric high-latitude temperature, leading to improved Equator-to-Pole temperature contrasts and zonal wind in both hemispheres. Though all models underestimate the Arctic winter near-surface temperature compared with the fifth generation of the European Center for Medium-Range Forecasts Reanalysis, validation against Arctic and Antarctic in-situ measurements reveals modest improvements in the refined experiments, with the measured values being better represented in the Arctic compared with the Antarctic and in summer compared with winter.

KEYWORDS

Arctic, Antarctic, global modeling, grid refinement, recent past, variable resolution

1 | INTRODUCTION

The climate in the polar regions is characterized by profound and rapid changes, both in the recent historical period and in future climate projections (e.g., Bintanja & Andry, 2017; Davy & Outten, 2020; Holland & Bitz, 2003; Rantanen *et al.*, 2022). Though mean surface temperatures over much of the Antarctic continent have not changed significantly, partly due to circulation changes associated with ozone depletion (e.g., Thompson *et al.*, 2011), new record maximum temperatures have been recorded twice in the Antarctic Peninsula in recent years (Francelino *et al.*, 2021). Offshore, Antarctic sea ice has lately undergone significant losses (e.g., Purich & Doddridge, 2023), with the last 4 years having the lowest four minima on record. In the Arctic, surface temperatures have been increasing at an accelerated rate compared with the global mean, a phenomenon known as Arctic amplification. Projections from the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring *et al.*, 2016) show that uncertainty in projected future warming is larger in the Arctic than anywhere else (Taylor *et al.*, 2022). This uncertainty arises from many sources, such as incomplete understanding of key processes, inaccurate representation of subgrid-scale processes, and interactions between these processes and the larger-scale climate.

Global climate models are excellent tools for investigating polar climate and its links to lower latitudes. However, the high computational cost renders it challenging to apply sufficiently high spatial and temporal resolution to properly represent important small-scale features like orographically induced gravity waves. Higher resolution over limited regions can be achieved at lower cost using regional climate models. However, such experiments rely on boundary conditions from reanalyses or global models that may lack full mathematical and physical consistency, and cannot simulate feedbacks to the large-scale global climate. Variable-resolution models, on the other hand, offer a way of employing high resolution in a limited area while using coarser resolution elsewhere, allowing two-way interactions between the high- and lower-resolution regions, reduced computational costs outside the refined region(s), and a unified set-up.

The development of variable-resolution models began in the 1970s with the stretched-grid models (e.g., Schmidt, 1977; Staniforth & Mitchell, 1978). Stretched grids are structured global grids in which the grid spacing is gradually increased away from the high-resolution region of interest. Results from an inter-comparison of four stretched-grid models with refinement over North America demonstrated smaller errors, particularly for orographically-induced precipitation (Fox-Rabinovitz

et al., 2008). Stretched-grid simulations with enhanced resolution over Antarctica exhibit improved representation of past and future Antarctic climate, including its surface mass balance (e.g., Krinner *et al.*, 1997, 2007, 2014). A drawback of stretched grids is that areas far away from the refined region are represented at very coarse resolution.

More recently, several unstructured-grid models have been developed, allowing for greater flexibility in the shape and position of the refined regions. The refinement is typically applied as a local refinement, as in the Community Earth System Model (CESM; Zarzycki *et al.*, 2014; Danabasoglu *et al.*, 2020), and the Model for Prediction Across Scales (MPAS; Skamarock *et al.*, 2012) or as a two-way nesting, as with the Geophysical Fluid Dynamics Laboratory dynamical core (Harris & Lin, 2013, 2014; Mouallem *et al.*, 2022) and in the Icosahedral Nonhydrostatic (ICON) model (Zängl *et al.*, 2015, 2022). Local refinement offers a smooth transition between coarse and refined regions, ameliorating many challenges associated with abrupt changes in two-way nesting, but also requiring scale-aware parametrizations and time steps determined by the finest resolution. In contrast, two-way nesting facilitates distinct time steps and physical-parametrization settings in different regions; a limitation of this approach is that spurious waves can be induced at the nest boundaries.

Both types of unstructured-grid models have been used for a wide range of regions, including high-mountain areas in Asia (e.g., Rahimi *et al.*, 2019; Wijngaard *et al.*, 2023), East Asia (e.g., Cheng *et al.*, 2023; Liang *et al.*, 2021; Wang *et al.*, 2024; Zhu *et al.*, 2022), North America (e.g., Gettelman *et al.*, 2018; Lin *et al.*, 2024; Rhoades *et al.*, 2018; Tang *et al.*, 2019; Wu *et al.*, 2017), North Atlantic (e.g., Wills *et al.*, 2024), South America (e.g., Bambach *et al.*, 2022), the Mediterranean region (e.g., Boza *et al.*, 2025), and Africa (e.g., Maoyi & Abiodun, 2022). All studies report a general reduction of climatological biases in many meteorological variables with refinement. For example, improvements are found for orographic precipitation and snow-water equivalent (e.g., Gettelman *et al.*, 2018; Tang *et al.*, 2019; Wang *et al.*, 2024), for characteristics of the East Asian monsoon (e.g., Liang *et al.*, 2021; Zhu *et al.*, 2022), albeit with no consensus for the monsoon precipitation evolution (e.g., Wang *et al.*, 2024), and for characteristics of tropical cyclones and associated large-scale circulation and precipitation patterns (e.g. Liang *et al.*, 2023; Zarzycki *et al.*, 2014).

Fewer studies have, however, targeted the polar regions. CESM experiments with refinement over the Arctic (Herrington *et al.*, 2022), Greenland (Herrington *et al.*, 2022; van Kampenhout *et al.*, 2019), and Antarctica (Datta *et al.*, 2023) revealed that the surface mass balance mainly improves with refinement whereas results are more mixed for clouds, with some degradations found in Datta *et al.* (2023) and van Kampenhout *et al.* (2019) and

improvements in Herrington *et al.* (2022). Furthermore, Waling *et al.* (2024) found the extent of atmospheric rivers and precipitation more closely matches reanalysis in CESM experiments with refinement over Greenland.

To further examine the effect of polar refinement on model performance, we carried out a coordinated set of global climate model experiments for the recent past (1985–2014) with refined regions centered over the Poles with the finest horizontal resolution poleward of 60° (the “refined” experiments). To evaluate the influence of the refinement, we also performed baseline experiments without refinement (the “unrefined” experiments), using the same coarse resolution everywhere. We use three state-of-the-art models: CESM, ICON, and MPAS. The experiments are “AMIP-style”; that is, in the style of the Atmosphere Model Intercomparison Project experiments (Eyring *et al.*, 2016; Gates *et al.*, 1999), with sea-surface temperatures (SSTs) and sea-ice concentrations (SICs) prescribed from forcing files while the other model components are fully active. We chose the AMIP-style approach to simplify the analysis and isolate atmospheric interactions; furthermore, the development of fully coupled variable-resolution models with refinement capability in the atmosphere, land, ocean, and sea-ice components only started recently (e.g., Huo *et al.*, 2025). To ensure comparability, we follow a common modeling protocol detailing key aspects of the model set-up. To the best of our knowledge, the result is the first coordinated intermodel comparison of global variable-resolution unstructured-grid model experiments for the polar regions.

The purpose of this article is to document the experiment protocol and the set-up of the models (Section 2) and present the first set of results for this multimodel ensemble (Section 4). We assess refinement-induced changes and their influence on model errors (with respect to the reference data in Section 3), focusing on temperature, zonal wind, and surface heat fluxes. Process-based studies further exploring mechanisms behind results reported here are currently ongoing. Indeed, Arctic temperature and precipitation extremes in the CESM experiments are already analyzed in Wijngaard *et al.* (2025). We summarize and discuss our results in Section 5.

2 | EXPERIMENTS AND MODELS

2.1 | Common experiment protocol

Our experiment set comprises global simulations for the recent past (1985–2014) with polar refinement (refined experiments) and without refinement; that

is, with uniform resolution (unrefined experiments). Both were carried out with three climate models with regional-refinement capabilities: CESM, ICON, and MPAS. Whereas CESM has a long history, dating back to the Community Climate Model first released 1983, ICON and MPAS were developed more recently. To ensure that the results are as comparable as possible, we configured the experiments following the same protocol, described in the following.

In the refined experiments, the grids were constructed with the equatorward edge of the finest-resolution region as close to 60° as technically feasible, with a transition region extending to around 40° in CESM, 50° in ICON and 30° in MPAS, and with coarse resolution equatorward of the transition zone (Figure 1). In the unrefined experiments, the grids have uniform resolution matching the coarsest used in the refined grids. ICON and MPAS were run with one refined region per grid, requiring three separate grids: (a) Arctic refined with finest resolution poleward of 60° N; (b) Antarctic refined with finest resolution poleward of 60° S; and (c) unrefined with the coarsest resolution everywhere. CESM was run with Arctic and Antarctic refinement simultaneously, requiring only two grids: (a) dual-polar refined with finest resolution poleward of 60° N and 60° S and (b) unrefined with the coarsest resolution everywhere. Altogether, we conducted eight experiments: the unrefined, Arctic-refined, and Antarctic-refined experiments with MPAS and ICON and the unrefined and dual-polar-refined experiments with CESM. The finest and coarsest horizontal resolutions correspond to ~10 km and ~50 km in ICON and MPAS and ~28 km and ~111 km in CESM due to technical and computational restrictions. See [Supporting Information Table S1](#) and Sections 2.2–2.4 for further details.

The experiments are AMIP-style, entailing that the atmosphere and land components were fully active, whereas the ocean and sea-ice components were replaced by data components. The SST and SIC fields were prescribed to daily values from the fifth generation of the European Center for Medium-Range Forecasts Reanalysis (ERA5; Hersbach *et al.*, 2020) for the years 1985–2014. Although SIC is the same in all models, intermodel differences can arise in and around the sea-ice regions due to differences in surface fluxes, sea-ice thickness (which is calculated by the sea-ice model in our ICON experiments, but prescribed to 2 m in the Northern Hemisphere [NH] and 1 m in the Southern Hemisphere [SH] in CESM and to 1 m in both hemispheres in MPAS), and so forth (Sections 2.2–2.4).

We initialized the experiments on January 1, 1984, using each model’s default method and spun them up for

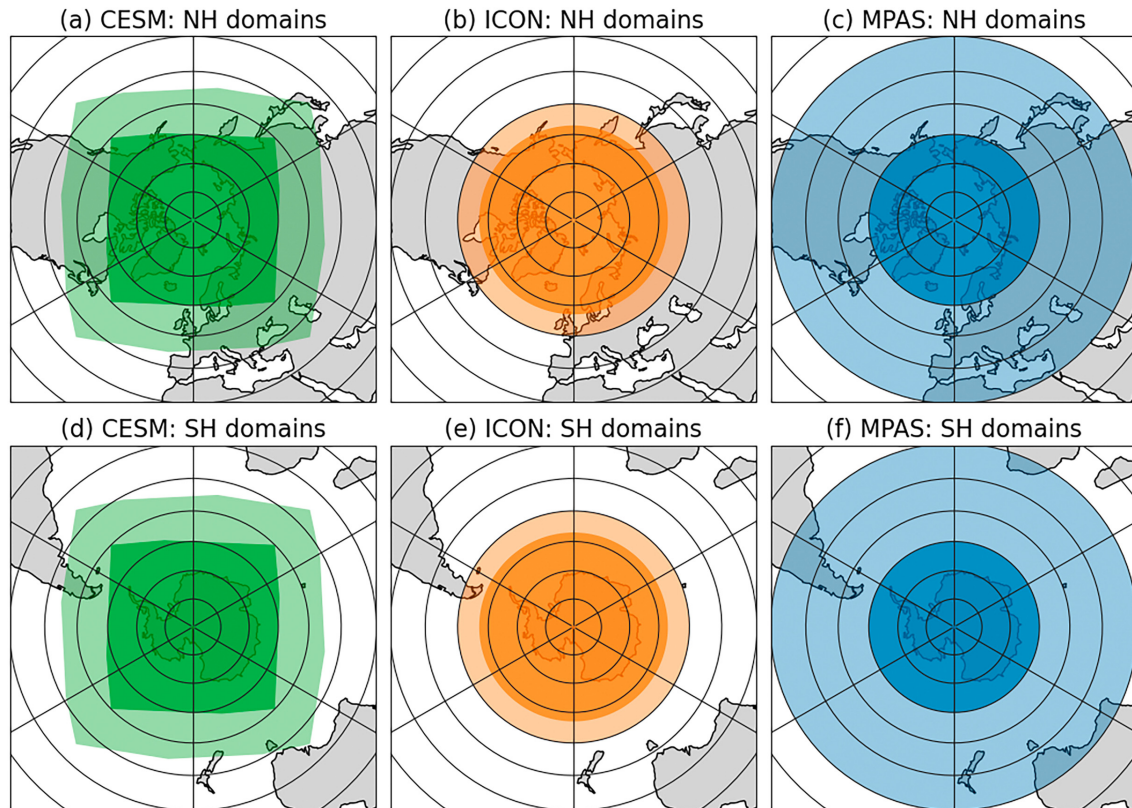


FIGURE 1 Overview of the horizontal extent of the regions with finest resolution (saturated colors) and intermediate/transitional resolution (lighter colors) in the refined grids used for the (a, d) Community Earth System Model (CESM; green regions), (b, e) Icosahedral Nonhydrostatic (ICON) model (orange regions), and (c, f) Model for Prediction Across Scales (MPAS; blue regions). Note that in CESM, the Northern Hemisphere (NH) and Southern Hemisphere (SH) domains are incorporated into the same grid (dual-polar refined), whereas we use separate grids for ICON and MPAS (Arctic refined and Antarctic refined; see Section 2). Though the transitional region is illustrated with a single hue in each panel, the grid spacing in CESM and MPAS gradually transitions from the finest to the coarsest resolution ([Supporting Information Table S1](#)).

12 months with one ensemble member for each experiment. We used time-varying forcings for greenhouse gases (GHGs: carbon dioxide, CO_2 ; methane, CH_4 ; nitrous oxide, N_2O ; and chlorofluorocarbons, CFCs), ozone, and tropospheric aerosols based on the CMIP6 historical forcings (Meinshausen *et al.*, 2017). In cases where the models do not allow for time-varying concentrations of certain species, we used climatologies representative for the recent past (Sections 2.2–2.4). Stratospheric aerosols were not included.

All models include a common set of variables in their output, comprising monthly and daily means and six-hourly instantaneous fields (see [Supporting Information Tables S5 and S6](#) for an overview). The output was interpolated to regular grids corresponding to each model's unrefined resolution as a post-processing step. In the vertical, the output was interpolated to the plev19 pressure levels defined in the CMIP6 data request (Jukes *et al.*, 2020, version 01.00.31).

2.2 | CESM

The CESM simulations were performed with version 2.2 (Danabasoglu *et al.*, 2020). The atmosphere component is the Community Atmosphere Model version 6 (CAM6), applied with a hydrostatic spectral element dynamical core (Gettelman *et al.*, 2019b; Lauritzen *et al.*, 2018) that supports unstructured grids without polar singularities and enables regional refinement (Zarzycki *et al.*, 2014). The CAM6 physical parametrizations include a shallow convection, boundary-layer turbulence, and cloud macrophysics scheme (Cloud Layers Unified by Binormals; Bogenschutz *et al.*, 2013), a deep convection scheme (Zhang & McFarlane, 1995), a cloud microphysics scheme with prognostic treatment of precipitation (Gettelman & Morrison, 2015), an orographic-drag parametrization scheme (Beljaars *et al.*, 2004), a modified modal aerosol module (Liu *et al.*, 2016), and a cloud–aerosol radiation scheme (Iacono *et al.*, 2008). The land component is the

Community Land Model version 5 (Lawrence *et al.*, 2019), applied with prescribed satellite vegetation phenology. Topography was interpolated from an updated 30-arcsec global topography dataset (Wijngaard *et al.*, 2023) consisting of the Global Multi-resolution Terrain Elevation Data (GMTED2010; Danielson & Gesch, 2011) and the Bed-Machine topography (Morlighem *et al.*, 2017, 2020) for Greenland and Antarctica. Whereas the SSTs and SICs were prescribed (Section 2.1), the Community Ice Code version 5 (Hunke *et al.*, 2015) calculated surface fluxes, snow depth, albedo, and surface temperature over ice. We prescribed sea-ice thickness to the standard CESM values for AMIP-style runs: 2 m in the NH and 1 m in the SH.

We calibrated short-wave radiation, sea-ice albedo, and low-level cloudiness following (Wijngaard *et al.*, 2023). Furthermore, we rectified a bug in the cloud microphysics by increasing the number of microphysical sub-steps and removing the limiter on the cloud ice-number concentrations (Shaw *et al.*, 2022; Zhu *et al.*, 2022). For the unrefined simulations, we used eight microphysical sub-steps, corresponding to a microphysical time step of 75 s, following Zhu *et al.* (2022). For the refined simulations, we maintained the microphysical time step of 75 s, corresponding to three sub-steps, which is computationally cheaper while performing similarly to test simulations with eight sub-steps.

We used two grids: the unrefined grid with 1° (~111 km) resolution everywhere and the dual-polar refined grid with 0.25° (~28 km) resolution poleward of 60° N and 60° S, a transitional region with 0.5° resolution between 60° N–45° N and 60° S–45° S to ensure numerical stability, and a coarse-resolution region with 1° resolution equatorward of 45° (Figure 1a,d and Supporting Information Table S1). We generated the dual-polar grid using the SQadGen software package (Ullrich, 2014). All experiments had 58 hybrid-sigma levels with the model top at ~40 km, resulting in finer vertical resolution than the 32-level grid used for CMIP6, particularly in the planetary boundary layer and upper troposphere (Danabasoglu *et al.*, 2020; Huang *et al.*, 2022; Lauritzen *et al.*, 2018). The computation cost was ~3,500 CPU hours per model year for the unrefined experiments and ~50,000 CPU hours per model year for the dual-polar refined experiments (Supporting Information Table S2).

The experiments were initialized from a 58-level CAM6 experiment representative of the recent past. Time-varying GHG forcings for CO₂, CH₄, N₂O, CFC-11, and CFC-12 were obtained from CESM2 experiments with the Whole Atmosphere Community Climate Model version 6 (WACCM6; Gettelman *et al.*, 2019a) forced in accordance with the CMIP6 protocol, in line with

the official CESM2–CAM6 CMIP6 experiments (Danabasoglu *et al.*, 2020); time-varying ozone concentrations and aerosols were taken from the same CESM2–WACCM6 experiments (Danabasoglu *et al.*, 2020; Hoesly *et al.*, 2018; van Marle *et al.*, 2017).

2.3 | ICON

ICON is a flexible, scalable, high-performance modeling framework for weather and climate applications. While the non-hydrostatic dynamical core (Zängl *et al.*, 2015) allows for applying the model on a wide range of scales, its icosahedral grid enables a high uniformity in the resolution over the whole sphere, thus avoiding polar singularities of latitude–longitude grids. We ran all experiments with version 2.6.6, configured with the multicomponent radiation scheme ecRad (Hogan & Bozzo, 2018), a single-moment cloud microphysics scheme based on Doms *et al.* (2011) and Seifert (2008), a Tiedtke–Bechtold-based convection scheme (Bechtold *et al.*, 2008; Tiedtke, 1989), a prognostic turbulent kinetic energy turbulence scheme (Raschendorfer, 2001), a subgrid-scale orographic drag scheme based on Lott and Miller (1997), a non-orographic gravity wave drag scheme based on Orr *et al.* (2010), the TERRA land component for the soil–vegetation–atmosphere-transfer based on Schrodin and Heise (2001), and topography from the GLOBE dataset with a raw resolution close to 1 km (Hastings *et al.*, 1999). Even though ICON was run with prescribed SSTs and SICs (Section 2.1), the sea-ice scheme still accounts for thermodynamic processes in order to calculate the sea-ice thickness using a bulk approach to solve the heat transfer equation (Mironov *et al.*, 2012).

ICON was run with three grids (Supporting Information Table S1). For the unrefined experiments, we applied the R3B5 resolution (~52.6 km) globally. We created the high-resolution domains in the Arctic- and Antarctic-refined experiments (Figure 1b,e) with two-way nesting, allowing the coarse- and intermediate-resolution domains to overlap with and exchange information with the high-resolution domain. As each nested subdomain can only have half of the grid spacing of the parent domain, two nested subdomains were needed to reach the desired resolution over the Poles: an intermediate domain north/south of 50° N/S with a resolution of R3B6 (~26.3 km), and a high-resolution domain north/south of 57° N/S with a resolution of R3B7 (~13.2 km). We ran the model with 90 vertical levels with the model top at 75 km in all cases. Per halving of the grid spacing, we also halved the time step for each subdomain. The physical parametrizations were consistent across domains and were chosen

to realistically represent the given range of resolutions. Parameters representing the set-up of the subgrid-scale orographic drag and the non-orographic gravity wave drag were based on Köhler *et al.* (2021). The computation cost was ~5,400 CPU hours per model year for the unrefined experiments and ~29,000 CPU hours per model year for the Arctic/Antarctic-refined experiments (Supporting Information Table S2).

We initialized the experiments from ERA5 data. GHGs were prescribed to the time-varying forcings from the CMIP6 historical experiment (Meinshausen *et al.*, 2017). The ozone concentrations were prescribed to climatologies based on a combination of the Geostationary Environment Monitoring Spectrometer and the Monitoring Atmospheric Composition and Climate Reanalysis (Hollingsworth *et al.*, 2008; Inness *et al.*, 2013, 2019), and aerosols were from the Tegen climatology (Tegen *et al.*, 1997).

2.4 | MPAS

The atmospheric component of MPAS was designed to form the core of a state-of-the-art numerical weather and climate prediction model capable of resolving regional climate variability at fine scales while retaining global coverage. MPAS employs a C-grid staggered solver for state-variable fluxes on an unstructured centroidal Voronoi mesh with variable tile sizes (Ringler *et al.*, 2010; Thuburn *et al.*, 2009). The grid cell structure enables smooth transitions from finer to coarser tile sizes, preempting computational issues arising from abrupt grid resolution transitions typical of nested grids. The dynamical core solves the fully compressible, non-hydrostatic equations of motion (Skamarock *et al.*, 2012). Vertical fluxes are resolved on a geometric-height coordinate, consisting of 55 vertical levels, using a split-explicit time-integration scheme that handles the gravity waves on the acoustic substep (Klemp *et al.*, 2007). A third-order Runge–Kutta time-integration scheme is used to resolve a wide range of atmospheric modes, from weather to acoustic modes (Wicker & Skamarock, 2002).

All experiments were carried out with MPAS version 7.3 using the default mesoscale physics suite, which is based on the Advanced Research Weather Research and Forecasting model physics. It includes the Tiedtke scheme for deep and shallow convection (Bechtold *et al.*, 2014), the six-class GRAUPEL phase microphysics scheme of the Single-Moment Micro-Physics for cloud nucleation (Hong & Lim, 2006), a vertical diffusion representation of non-local turbulent mixing in the boundary layer (Hong & Pan, 1996), and a representation of mountain-induced gravity wave drag from subgrid-scale orography (Choi &

Hong, 2015). Short-wave and long-wave radiative fluxes in the atmosphere and at the surface were computed by the Rapid Radiative Transfer Model for General Circulation Models (Iacono *et al.*, 2008), which accounts for natural and anthropogenic aerosols. Near-surface fluxes were computed using a Monin–Obukhov flux-profile scheme (Jiménez *et al.*, 2012), whereas surface fluxes over terrestrial areas were computed by the community Noah land surface model (Chen *et al.*, 1996; Chen & Dudhia, 2001; Ek *et al.*, 2003). Topography was taken from GMTED2010 and interpolated to the meshes used here. MPAS was run with prescribed SSTs and SICs (Section 2.1), with the sea-ice thickness prescribed to a uniform value of 1 m and the heat capacity and conductivity of the sea ice being set accordingly.

Similar to ICON, MPAS was run with three grids (Supporting Information Table S1). The unrefined experiment used a uniform 48 km resolution mesh (256,002 horizontal grid cells). The Arctic- and Antarctic-refined experiments used a 46 km–12 km variable-resolution mesh (655,362 horizontal grid cells) with the finest resolution (12 km) centered around the geographic North Pole and South Pole and extending equatorward to 60° N and 60° S respectively. Outside the refined region, tiles progressively became wider between 60°N/S and 30°N/S, transitioning to the coarse 46 km resolution used over the rest of the globe (Figure 1c,f). The computational cost of the unrefined and Arctic/Antarctic-refined experiments was ~50,000 and ~350,000 CPU hours per model year (Supporting Information Table S2).

CO₂ was prescribed to time-varying annual-mean global-mean values from the CMIP6 historical forcings. Other GHGs (CH₄, N₂O, CFCs, and carbon tetrachloride) were prescribed to the global-mean and annual-mean concentrations from 2005. Aerosols and ozone concentrations were prescribed to monthly and daily climatologies respectively.

3 | REFERENCE DATA

We compare the models' data with reference data from ERA5 (Hersbach *et al.*, 2020) for near-surface temperature, (three-dimensional) temperature, zonal wind, near-surface wind, near-surface specific humidity (derived from dew-point temperature and surface pressure fields), cloud cover, and sea-level pressure. In addition, we compare the models' near-surface temperature with in-situ measurements from 10 meteorological stations in each polar region (locations shown in Supporting Information Figure S1); these stations were chosen to have a widespread spatial distribution and also near-complete monthly data for 1985–2014. We evaluate

surface latent and sensible heat fluxes against FLUX-COM (Jung *et al.*, 2019; Tramontana *et al.*, 2016) over land and the Woods Hole Oceanographic Institution (WHOI) Objectively Analyzed Air–Sea Fluxes (OAFux; Yu *et al.*, 2008) version 3 over ocean. For the surface downwelling and upwelling long-wave (rlds and rlus) and short-wave (rsds and rsus) fluxes (abbreviated variable names are given in lower case, consistent with the CMIP6 variable names and the Climate Model's Output Rewriter standards), we use monthly mean surface fields from version 4.2 of the Clouds and the Earth's Radiant Energy System Energy Balanced and Filled (CERES-EBAF) dataset (NASA/LARC/SD/ASDC, 2023). In [Supporting Information](#) Figures S19–S22, we additionally consider reference data from the Japanese Reanalysis for Three Quarters of a Century (JRA-3Q; Kosaka *et al.*, 2024) for temperature, zonal wind, and sea-level pressure along with surface heat fluxes from ERA5. We use years 1985–2014 for all datasets except CERES-EBAF, where we use 2001–2025.

4 | RESULTS

In what follows, we examine the differences between the refined and unrefined experiments to assess whether refinement improves model errors. We first focus on temperature, zonal wind, and surface fluxes of latent and sensible heat for the NH and SH extratropics, defined as latitudes poleward of 20°. Then we zoom in on the polar regions, evaluating the effect of the refinement across a more comprehensive set of variables to identify processes that warrant further investigation. We focus on the respective winter season, December–February (DJF) for the NH and June–August (JJA) for the SH, when the atmosphere tends to be more dynamically active. However, results for the summer seasons (JJA for the NH and DJF for the SH) are included in the [Supporting Information](#) and discussed briefly later herein. In line with our focus being on winter, the results presented in the following pertain to this season unless stated otherwise. For ICON and MPAS, we consistently use the Arctic-refined runs for NH and Arctic results, and the Antarctic-refined runs for SH and Antarctic results. For CESM, we use the same (dual-polar) refined run in all cases. All model results are based on years 1985–2014.

Although the refined grids achieve their finest resolution in the polar regions, horizontal resolution is also enhanced at lower latitudes due to the transition zone between the finest and coarsest resolution (Section 2, Figure 1, [Supporting Information](#) Table S1). In contrast to the finest-resolution region, the latitudinal extent of the

transition zone varies considerably between the models (Figure 1): MPAS has the largest, extending from 60° to 30°, whereas ICON has the smallest, extending from 57° to 50°.

4.1 | Temperature

As expected in AMIP-style experiments, NH winter near-surface temperature biases are small over open oceans and larger over land and ice-covered regions in the unrefined experiments (Figure 2a–c). A common feature is the cold bias over the Arctic Ocean and Greenland, which is stronger and more extensive in ICON and MPAS, extending throughout the entire North American continent and most of northern Eurasia (Figure 2b,c). These cold biases are associated with underestimated rlds (Supporting Information Figure S2a,e,i), particularly in ICON and MPAS where cloudiness is underestimated in the same regions (Supporting Information Figure S2g,k). The high-latitude near-surface temperature biases should be interpreted with some caution, as they partly reflect ERA5 being too warm over sea ice in winter (e.g., Batrak & Müller, 2019; Tian *et al.*, 2024; Zampieri *et al.*, 2023); to better assess the magnitude these biases, we therefore compare the near-surface temperature from the models, and from ERA5, with observations from 10 Arctic and Antarctic surface stations in Section 4.1.2.

At lower latitudes, the unrefined CESM run is too warm over southern North America and large parts of Eurasia (Figure 2a). Warm biases are also present over the North Atlantic and the eastern North Pacific. These biases, however, have a relatively limited vertical extent (Figure 3a). This, combined with the consistent positive biases in rlds and cloud cover (Supporting Information Figure S2a,c), suggests that they may be linked to the new vertical scheme in CESM's atmospheric component CAM, which has enhanced vertical resolution in the planetary boundary layer. ICON also overestimates the cloud cover over the NH oceans, which, together with negative rlds biases and small near-surface temperature biases, suggests cloud property errors (Figure 2b, Supporting Information Figure S2e,g). MPAS, on the other hand, underestimates cloudiness, with consistent biases in rlds and rsus (Supporting Information Figure S2i,k; rsus not shown). Overall, CESM has the smallest errors in near-surface temperature and MPAS the largest, both in terms of the area-mean bias and root-mean-square error (RMSE; left and right titles above Figure 2a–c).

The refinement-induced changes in NH near-surface temperature are characterized by high-latitude cooling (Figure 2g–i), resulting in even stronger cold biases than

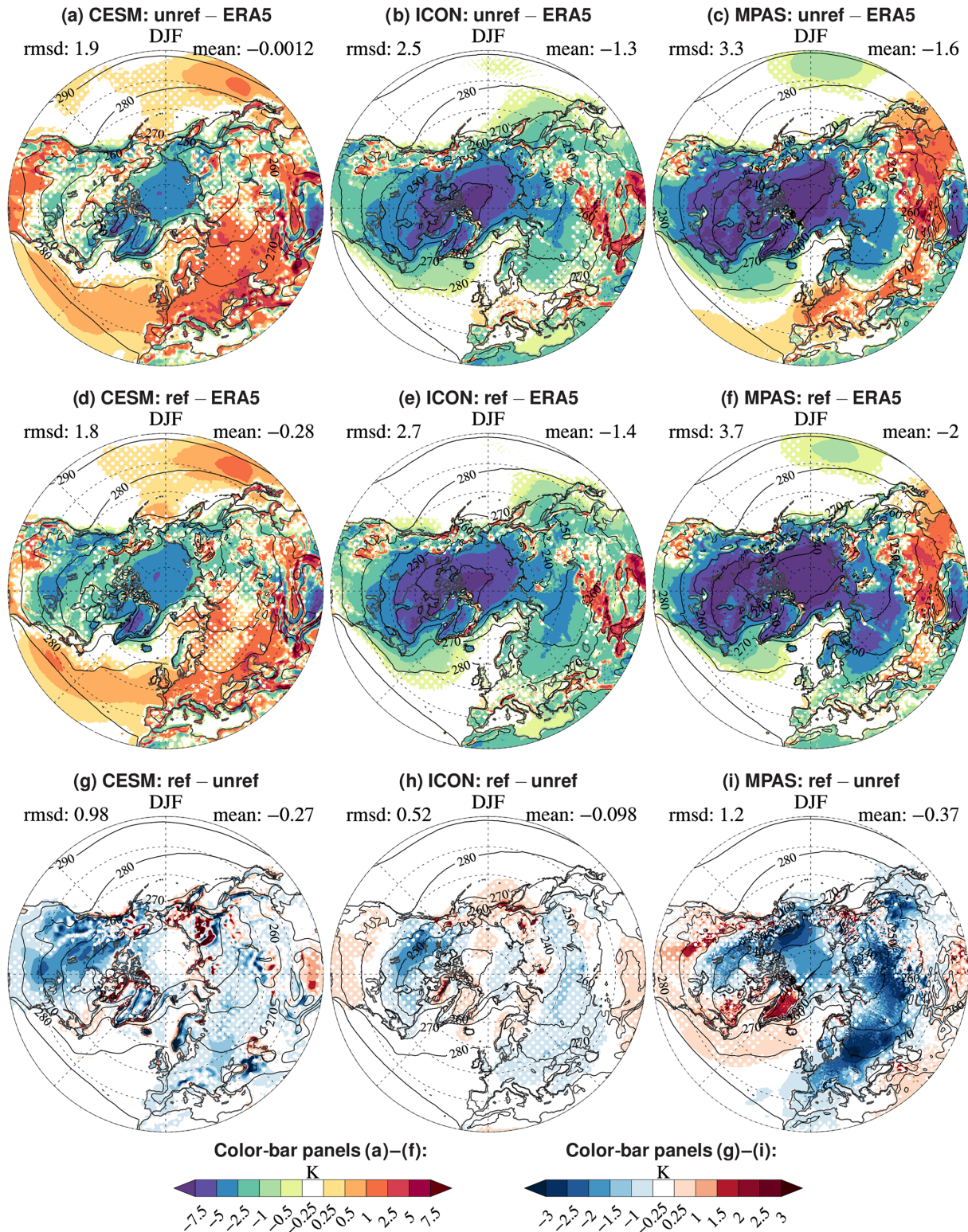


FIGURE 2 Northern Hemisphere winter (December–February, DJF) near-surface temperature biases (left color bar) for the (a)–(c) unrefined and (d)–(f) refined experiments and (g)–(i) the difference between the refined and unrefined experiments (right color bar) for (a, d, g) the Community Earth System Model (CESM), (b, e, h) the Icosahedral Nonhydrostatic (ICON) model, and (c, f, i) the Model for Prediction Across Scales (MPAS). The biases (a)–(f) are computed with respect to the fifth generation of the European Center for Medium-Range Forecasts Reanalysis (ERA5). In each panel, the area mean and root-mean-square-difference (rmsd) of the bias/difference field (colors) are given in the upper right and left corners. The solid black lines are selected contours from the respective unrefined climatology for reference. White dots indicate changes that are not significant at the 95% level according to a two-sided t test. Units are kelvin.

in the unrefined runs (relative to ERA5; comparing Figure 2d–f with a–c). The cooling is most pronounced in MPAS, particularly over northern Eurasia, the Arctic Ocean, and parts of northern North America, and generally weaker in CESM and ICON, with changes in ICON mostly being insignificant. The temperature changes tend to be accompanied by similar changes in rlds (Supporting Information Figure S2b,f,j), and largely also in cloud cover (Supporting Information Figure S2d,h,l). The area-mean NH near-surface temperature bias and RMSE are worse in the refined runs, except in CESM where RMSE improves slightly while the bias is larger (the latter indicating less efficient cancellation between positive and negative errors).

Though the NH high-latitude near-surface temperature biases do not generally improve with refinement, there are still clear regional improvements. In CESM, the warm bias over Europe and parts of western Asia is improved by the cooling in the same regions in the refined run, related to decreased rlds and cloud cover (Figure 2a,d,g and Supporting Information Figure S2b,d). In ICON, the cold bias over the northernmost part of the North Pacific and the adjacent continents is improved by the local patches of warming in the refined runs (Figure 2b,e,h). The warming over southern North America and the western North Atlantic in MPAS improves the cold bias in these regions, associated with increased rlds in the latter case; also, the cooling over southern Europe and central Asia, associated with reduced rlds, acts to ameliorate the warm biases in these regions (Figure 2c,f,i and Supporting Information Figure S2j).

The NH zonal-mean temperature results reveal that the high-latitude cold bias is confined to the lowest tropospheric levels in the unrefined MPAS and CESM experiments, whereas it extends throughout the troposphere in ICON (Figure 3a–c). All models are, moreover, too cold above ~500 hPa in the Arctic, and over large parts of the tropical troposphere and lower stratosphere. Too low upper-level temperatures is a longstanding problem in climate modeling (Hamilton *et al.*, 1995), often related to too strong radiative cooling in the upper troposphere and lower stratosphere due to an overly moist tropopause (Bland *et al.*, 2021) and too weak dynamical warming at higher levels in the stratosphere associated with poorly represented dynamical variability (Charlton-Perez *et al.*, 2013; Lawrence *et al.*, 2022). The largest intermodel disparity occurs in the Arctic between 950 and 500 hPa, where MPAS and CESM are too warm and ICON is too cold (Figure 3a–c).

The most striking change in NH zonal-mean temperature occurs in MPAS, where the Arctic upper troposphere and lower stratosphere warm substantially with

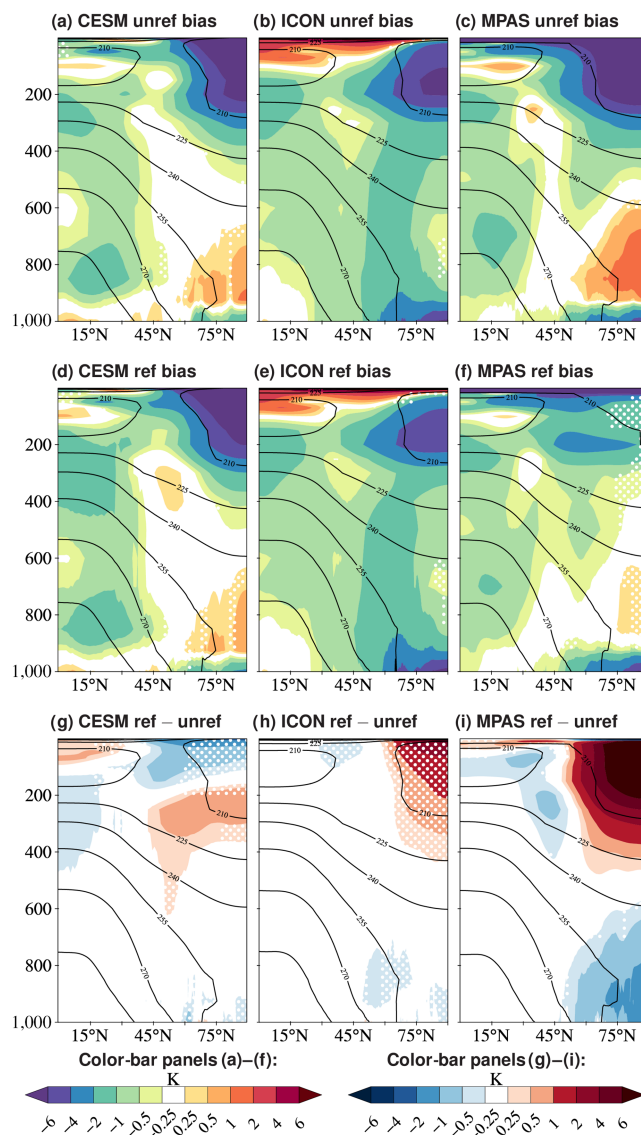


FIGURE 3 Northern Hemisphere winter (December–February) zonal-mean temperature biases (left color bar) for the (a)–(c) unrefined and (d)–(f) refined experiments and (g)–(i) the difference between the refined and unrefined experiments (right color bar) for (a, d, g) the Community Earth System Model (CESM), (b, e, h) the Icosahedral Nonhydrostatic (ICON) model, and (c, f, i) the Model for Prediction Across Scales (MPAS). Biases (a)–(f) are computed with respect to the fifth generation of the European Center for Medium-Range Forecasts Reanalysis. Black lines and white dots are as in Figure 2. Units are kelvin.

refinement (Figure 2i). This results in a pronounced improvement of upper-level Arctic temperature, given that this region is too cold in MPAS in the unrefined run (Figure 3c,f,i). At lower levels, MPAS cools with refinement, improving the bias between 950 and 600 hPa, where the model is too warm in the unrefined run, while worsening it below 950 hPa, where the model is too cold (Figure 3c,f,i). Overall, the representation of

NH zonal-mean temperature improves considerably above 950 hPa in MPAS. In ICON and CESM, the changes appear weak and insignificant in the zonal mean (Figure 3g,h), except between 400 and 200 hPa where CESM warms with refinement, improving the cold bias there (Figure 3a,d,g). However, Section 4.4 will show that the Arctic-mean RMSE clearly improves for all models when considering the 250 hPa temperature.

In the SH, results for near-surface temperature are largely consistent with the NH: during winter, biases in the unrefined runs are generally small over the ocean-covered regions and larger over the Antarctic continent and surrounding sea ice, where all models are too cold (compared with ERA5; Figure 4a–c). Again, we find the largest errors in MPAS, which is colder over and surrounding Antarctica than the other models, and this is also reflected in the area-mean bias and RMSE. Over the sea-ice-covered regions, the cold bias is linked to MPAS having a stronger negative rlds bias; however, biases over the Antarctic continent are less pronounced (Figure 4c and Supporting Information Figure S3i). The negative Antarctic near-surface temperature and rlds biases are not clearly linked to biases in cloud cover (Figure 4a–c, Supporting Information Figure S3a,c,e,i and c,g,k), suggesting that other processes, such as problems with the turbulence parametrizations under stable stratification, could play a role. In ICON, the cloud cover is overestimated between 20° S and 40° S, in a region where rlds is too low, suggesting cloud property errors. Similar to the NH, CESM exhibits warm biases over the Southern Ocean with consistent positive biases in rlds and cloud cover (Figure 4a, Supporting Information Figure S3a,c) and a negative rsds bias (not shown). Interestingly, the warming maxima exceed 0.5 K in locations that coincide with the ridging of the zonal wave number 3 pattern (Goyal *et al.*, 2021, their Figure 1), over the western South Atlantic, central Indian Ocean, and south of New Zealand (Figure 4a).

The refinement-induced changes in SH near-surface temperature are most pronounced for MPAS, whereas changes are weak for ICON and CESM (Figure 4g–i). In ICON, the refined experiment is slightly colder, strengthening the area-mean RMSE and bias (Figure 4b,e,h). In CESM, the refinement causes patches of warming and cooling over and around Antarctic, possibly attributable to better resolved Antarctic topography, resulting in improved area-mean RMSE (Figure 4a,d,g). The refinement is associated with a small area of cooling downstream of the southern tip of the Antarctic Peninsula in all models, most pronounced in MPAS (Figure 4g–i). However, the most prominent response in MPAS is a warming over most of the Antarctic continent, contrary to what was seen for the NH, where the polar regions are predominantly ocean covered (Figures 2i and 4i); as a result, the

regional biases are improved in the Antarctic along with the area-mean bias and RMSE (Figure 4c,f). The cold bias over the sea-ice-covered regions in MPAS does not, however, improve. Changes in rlds are consistent with those in near-surface temperature over the Antarctic continent, whereas changes in cloud cover are not clearly linked (Figure 4g–i, Supporting Information Figure S3b,f,j and d,h,l).

The SH zonal-mean winter temperature plots (Figure 5a–c) show that the CESM midlatitude oceanic warm bias is relatively confined to the lowest levels (as in the NH; Figure 3a) and that all models are too cold in the Antarctic troposphere and stratosphere with minima near the surface and above 200 hPa. Lower-latitude biases are weaker and vary more between models. Antarctic refinement causes pronounced warming in MPAS that extends throughout the troposphere, with the strongest warming above 200 hPa (Figure 5i). Similar but weaker warming is also seen in CESM (Figure 5g). ICON displays warming above ~300 hPa in the Antarctic, but changes are generally small below (Figure 5h). As the unrefined runs are too cold in the Antarctic, particularly above ~200 hPa and below ~700 hPa, refinement improves the Antarctic zonal-mean temperature in all models (Figure 5).

During summer, the near-surface temperature biases are generally smaller than during winter and vary geographically between models in both hemispheres (Supporting Information Figures S4 and S6). Refinement causes regional changes but the overall performance is mostly unaffected, except in the SH for ICON and CESM, where a warm bias over the Antarctic continent (inverse in sign to winter; Figure 4) is slightly alleviated (Supporting Information Figure S6). As in winter, the upper-air biases are characterized by a too cold tropopause in all models and in both hemispheres (Supporting Information Figures S5 and S7) whereas other biases are more model dependent. Refinement alleviates the cold tropopause bias in all models, except in ICON in the NH, where the bias was relatively small in the unrefined run; in the free troposphere, the refinement increases the NH warm biases in CESM and the NH and SH warm bias in MPAS but improves the cold NH biases in ICON (Supporting Information Figures S5 and S7).

4.1.1 | Polar trends

Although the models underestimate climatological Arctic near-surface temperatures compared with ERA5 during winter, the annual trends are better represented. Arctic-mean annual-mean near-surface

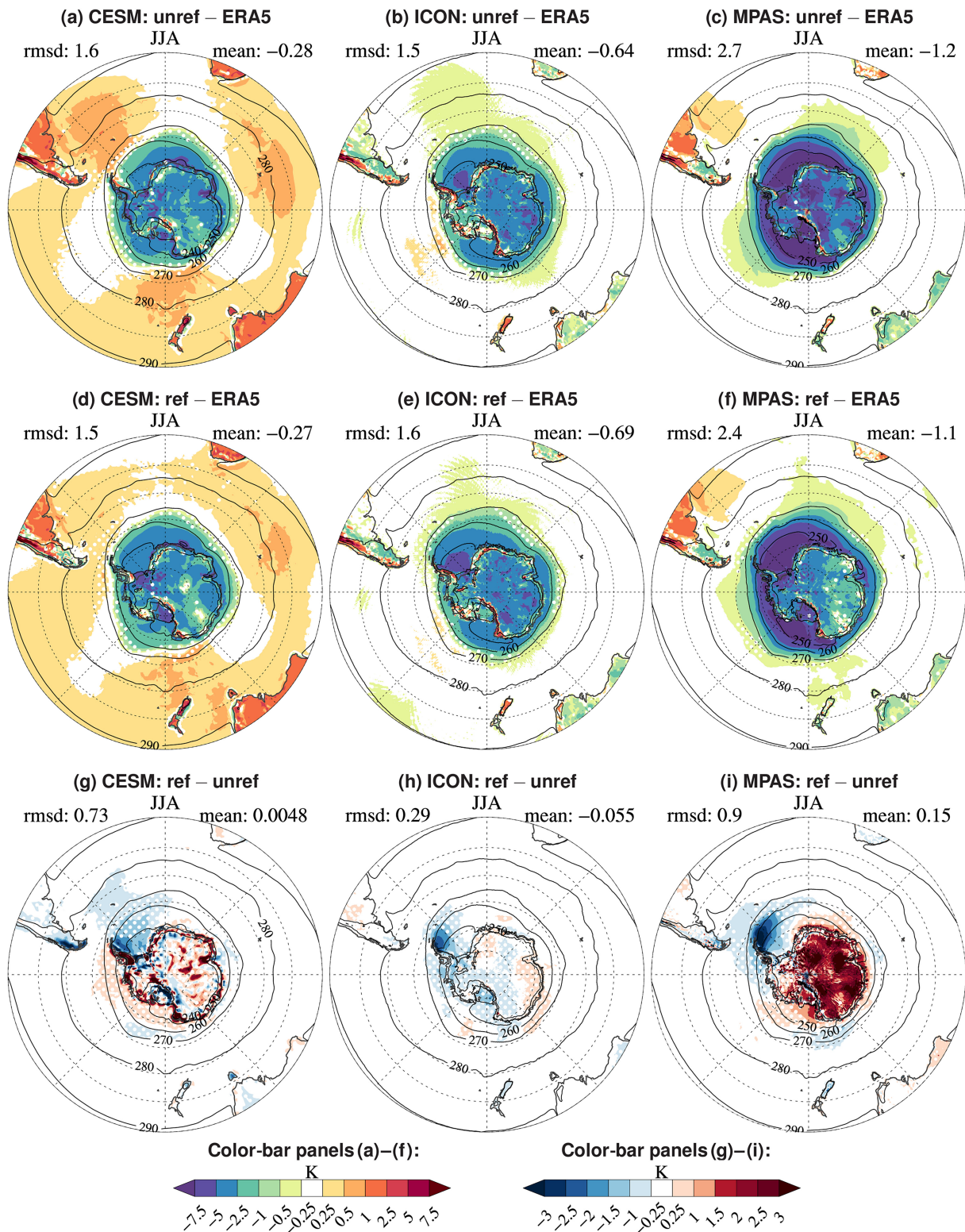


FIGURE 4 As in Figure 2, but for Southern Hemisphere winter (June–August, JJA) near-surface temperature. Units are kelvin.

temperature trends for 1985–2014 are $0.70 \text{ K} \cdot \text{decade}^{-1}$, $0.74 \text{ K} \cdot \text{decade}^{-1}$, and $0.62 \text{ K} \cdot \text{decade}^{-1}$ in the unrefined CESM, ICON, and MPAS experiments respectively, similar to the ERA5 trend of $0.69 \text{ K} \cdot \text{decade}^{-1}$ (Table 1).

With refinement, the CESM trend is unchanged, whereas the trends in ICON ($0.64 \text{ K} \cdot \text{decade}^{-1}$) and MPAS ($0.57 \text{ K} \cdot \text{decade}^{-1}$) weaken. Compared with ERA5, ICON somewhat overestimates the trend in the unrefined run

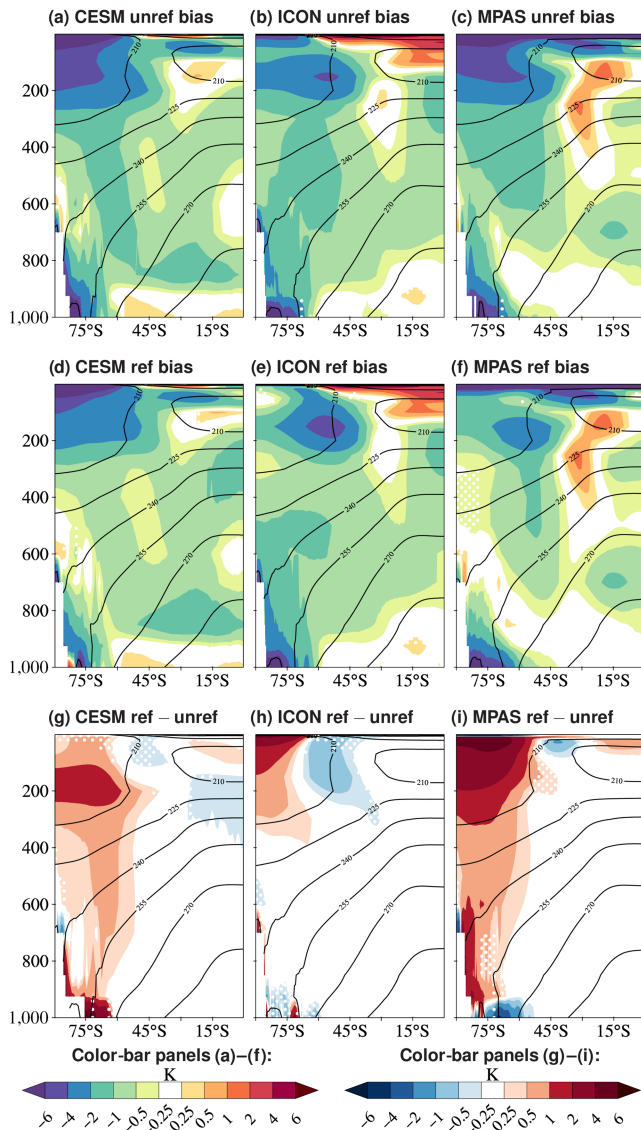


FIGURE 5 As in Figure 3, but for Southern Hemisphere winter (June–August) zonal-mean temperature. Units are kelvin.

and underestimates it in the refined run, whereas MPAS underestimates the trend in both cases.

In the Antarctic, the trends are negative and substantially weaker than in the Arctic. Compared with ERA5, where the trend is almost negligible and insignificant ($-0.03 \text{ K} \cdot \text{decade}^{-1}$), the trends are too strong in CESM, ICON, and MPAS in the unrefined runs (-0.12 , -0.16 , and $-0.19 \text{ K} \cdot \text{decade}^{-1}$). The effect of the refinement is mixed, with no changes in CESM, very small changes in ICON, and a stronger trend in MPAS (Table 1). We note that the recent near-surface temperature trends in ERA5 are known to be too large in certain regions of Antarctica (Bromwich *et al.*, 2024).

Exploring the effect of internal variability on the trends, we also consider ensemble-mean trends and spread from 10 members (r1i1p1f1–r10i1p1f1) from the CMIP6

CESM2 AMIP experiment (Danabasoglu, 2019; Danabasoglu *et al.*, 2020; Eyring *et al.*, 2016). The spread is $0.05 \text{ K} \cdot \text{decade}^{-1}$ for the Arctic and Antarctic trends, indicating that the influence of the refinement exceeds internal variability in the Arctic for ICON and in the Antarctic for MPAS. Though the Antarctic trends for ERA5 and for refined and unrefined CESM runs fall within the CMIP6 CESM2 AMIP range, the CMIP6 CESM2 AMIP trend is weaker relative to the other models and ERA5 in the Arctic and relative to ICON and MPAS in the Antarctic.

4.1.2 | Comparison with in-situ measurements

Although the comparison with ERA5 indicates that the models are too cold in the Arctic, recent literature finds ERA5 to be too warm in this region (Batrak and Müller, 2019; Zampieri *et al.*, 2023; Tian *et al.*, 2024), indicating that the cold biases could be exaggerated. To investigate this further, we validate the near-surface temperature from the models and ERA5 against observations from selected Arctic and Antarctic meteorological stations for winter and summer (Section 3). We note that ERA5 assimilates data from these stations.

Validation summary statistics (mean bias and RMSE across the 10 stations and 30 years) are provided in Table 2 and the locations of the stations in Supporting Information Figure S1. The models are largely more accurate in the Arctic than Antarctic, the exception being MPAS in summer, and they do better in summer than winter. Moreover, in the majority of examples, the refined runs do better than the unrefined ones, although the improvements are generally small. Generally, ICON is the best model (italic numbers) and MPAS the worst (bold numbers), with the exception that the unrefined CESM run has the smallest bias and refined CESM run has the smallest RMSE for Arctic winter.

Interestingly, Table 2 reveals that both the refined and unrefined ICON runs outperform ERA5 in Antarctic summer. Analysis of the individual station data indicates that ERA5 does especially badly at Vostok, situated on the Antarctic Plateau, where it has the second worst bias (-6.89°C) across all the models (not shown). This can be associated with the intense near-surface temperature inversions that reside over the continental interior (Xie *et al.*, 2023). McMurdo Station, located on the Ross Ice Shelf, has the worst statistics for most models in both winter and summer. The biases are always negative, implying that the models are failing to capture the protection that the Transantarctic Mountains provide McMurdo from the coldest plateau temperatures in winter and the influence of nearby open water in the summer, when the Ross

TABLE 1 Trends in near-surface temperature for the Arctic ($\geq 60^\circ$ N) and Antarctic ($\geq 60^\circ$ S) computed for the fifth generation of the European Center for Medium-Range Forecasts Reanalysis (ERA5) and the unrefined (U) and refined (R) experiments with the Community Earth System Model (CESM), the Icosahedral Nonhydrostatic (ICON) model, and the Model for Prediction Across Scales (MPAS).

	Polar near-surface temperature trend ($\text{K} \cdot \text{decade}^{-1}$)							
	ERA5	CESM-U	CESM-R	ICON-U	ICON-R	MPAS-U	MPAS-R	CESM2
Arctic	0.69*	0.70*	0.70*	0.74*	0.64*	0.62*	0.57*	0.45 ± 0.05 *
Antarctic	−0.03	−0.12*	−0.12*	−0.16*	−0.15	−0.19*	−0.29*	$−0.08 \pm 0.05$ *

Note: To assess the influence of internal variability, we also include the ensemble-mean trend and spread (± 1 standard deviation) computed from 10 realizations from the Coupled Model Intercomparison Project Phase 6 CESM2 Atmosphere Model Intercomparison Project experiment (Danabasoglu, 2019). Trends are based on a least-squares approach and marked with an asterisk when they are significant at the $P = 0.9$ level according to the Mann–Kendall non-parametric test. For CESM2, the significance is assessed for each member and the asterisk indicates significant trends for all 10 members. We use yearly means for 1985–2014 and the data were interpolated to a common grid as a preprocessing step to ensure that the same latitude thresholds can be used in all cases.

TABLE 2 Validation statistics in terms of mean bias and root-mean-square error (RMSE) for near-surface temperature based on in-situ observations from 10 Arctic and 10 Antarctic stations (locations shown in Supporting Information Figure S1) for the fifth generation of the European Center for Medium-Range Forecasts Reanalysis (ERA5) and the unrefined (U) and refined (R) experiments with the Community Earth System Model (CESM), the Icosahedral Nonhydrostatic (ICON) model, and the Model for Prediction Across Scales (MPAS) for years 1985–2014.

Statistic	ERA5	CESM-U	CESM-R	ICON-U	ICON-R	MPAS-U	MPAS-R
Arctic winter (DJF)							
Mean bias (K)	−0.29	−0.53	−1.33	−2.84	−1.46	−2.92	−3.58
RMSE (K)	1.44	3.04	3.02	3.42	3.17	10.95	11.18
Arctic summer (JJA)							
Mean bias (K)	−0.20	−2.40	−2.31	−0.84	−0.51	−4.39	−3.82
RMSE (K)	1.07	3.02	2.97	1.57	1.24	5.06	4.62
Antarctic winter (JJA)							
Mean bias (K)	−0.24	−4.63	3.55	−2.72	−2.72	−5.32	−4.72
RMSE (K)	3.07	6.49	5.15	4.14	4.51	11.27	10.74
Antarctic summer (DJF)							
Mean bias (K)	−1.63	−3.12	−2.58	−1.22	−0.99	−2.80	−2.12
RMSE (K)	2.72	4.19	3.47	2.20	1.80	7.12	6.70

Note: The model data are linearly interpolated to the location of the meteorological stations to 0.01° . For each region, season, and statistic, the value for the best model is in italics and the value for the worst is shown in bold.

Abbreviation: DJF: December–February; JJA: July–August.

Sea sea ice melts. There are marked differences in the statistics between the models at Vernadsky on the western side of the Antarctic Peninsula in winter. Local temperatures in this season are predominantly controlled by the offshore sea-ice extent in the Bellingshausen Sea. Given that sea-ice extent is prescribed to the same values in all models, other effects must be at play, such as inconsistencies in sea-ice thickness or snow on sea ice (Section 2). Other features in the models, such as the representation of orography or advection schemes (Sections 2.2–2.4) could also play a role. Nonetheless, the strong negative biases in the CESM and MPAS models at Vernadsky ($−10.43^\circ\text{C}$

in MPAS-R) suggest that key aspects of the local climate related to sea ice are not accurately captured. In contrast, the bias of only $−0.45^\circ\text{C}$ in ICON-R implies that the impact of sea-ice variability is better represented here.

In the Arctic, there are also some very large biases at individual stations. The greatest are in the refined MPAS run, ranging from $−22.40^\circ\text{C}$ at Jan Mayen to $+16.69^\circ\text{C}$ at Verhojansk. The CESM and ICON models do very well at Jan Mayen, an island where temperature is dominated by regional ocean currents and which has an annual temperature range $<10^\circ\text{C}$, so this issue is specific to the MPAS model. In contrast, the refined MPAS run has temperatures

that are far too high at Verhojansk, a location famous for being one of the coldest places on Earth in winter due to its extremely continental climate dominated by the climatological Siberian High high-pressure system. In MPAS, this circulation regime is not sufficiently developed to reproduce the very low winter temperatures ($\sim -45^{\circ}\text{C}$).

4.1.3 | Equator-to-Pole temperature contrasts

The zonal-mean temperature plots (Figures 3 and 5, Supporting Information Figures S5 and S7) reveal substantial improvements of the polar cold biases in the upper troposphere and lower stratosphere with refinement. Given the implications for the meridional temperature gradients, and, by thermal wind balance, for zonal wind, we explore this further by analyzing the Equator-to-Pole temperature contrasts at upper and lower levels (e.g., Graff *et al.*, 2019; Harvey *et al.*, 2014; Li *et al.*, 2018) in Figure 6.

During winter, we find the most pronounced and consistent improvements at 250 hPa (Figure 6a,b), in line with the zonal-mean temperature results. The temperature contrasts, which are too strong in the unrefined runs in all models and both hemispheres, weaken and are much more similar to the reanalyses in the refined runs. The improvement is particularly pronounced for MPAS (blue boxes) in both hemispheres and for CESM (green boxes) in the SH, where the interquartile range (IQR) for the refined runs lies entirely below that of the unrefined runs. CESM and ICON (orange boxes) also improve notably in the NH and SH respectively, with the median for the refined runs placed below the IQR for the unrefined run in both cases.

Results are more mixed for the low-level winter contrasts (850 hPa in the NH and 700 hPa in the SH; Figure 6c,d). In the NH, MPAS and CESM underestimate the contrasts, whereas ICON overestimates them; significant improvements are only found for MPAS, where the temperature contrasts are stronger and closer to the reanalyses in the refined experiment (non-overlapping IQRs; Figure 6c). In the SH, all runs underestimate the contrasts, with the refined runs being slightly worse than the unrefined for MPAS and CESM. Though both ERA5 and JRA-3Q (black boxes) feature stronger low-level SH contrasts than the models, disagreement between the reanalysis products indicates that the extent to which the models underestimate these contrasts is uncertain.

During summer, the low-level SH temperature contrasts in ERA5 and JRA-3Q demonstrate better agreement (Supporting Information Figure S8d). As in winter, the models generally underestimate these gradients in both the refined and unrefined runs; however, refinement reduces the spread in the MPAS run. The upper-level

temperature contrasts improve also during this season (Supporting Information Figure S8a,b) whereas low-level NH results are more varied (Supporting Information Figure S8c).

4.2 | Zonal wind

Changes in Equator-to-Pole temperature contrasts have been tied to changes in the storm tracks, both in fully coupled (e.g., Harvey *et al.*, 2014) and simpler models (e.g., Butler *et al.*, 2010). Harvey *et al.* (2014) found storm-track changes to be linked to both the upper- and lower-level gradients, particularly in the SH during both winter and summer, and to a lesser extent in the NH during winter. Given the marked improvement of the upper-level temperature contrast in Figure 6, we next explore whether the eddy-driven jet, here represented by 850 hPa zonal wind, changes accordingly.

In the NH, the 850 hPa zonal wind bias patterns are largely similar across the models in the unrefined runs (Figure 7a–c). The largest biases occur in the North Atlantic exit region, where the jet is too strong and too extended over Europe; this is particularly pronounced in MPAS, which has the largest errors overall. All models underestimate the North Atlantic jet on the equatorward side, and MPAS and ICON also underestimate it on the poleward side, over the Nordic Seas. Errors are smaller over the North Pacific, where MPAS underestimates the jet and CESM and ICON places it too far north.

Refinement considerably improves the 850 hPa North Atlantic jet in MPAS (Figure 7c,f,i). Similar changes are also found in ICON, but with weaker amplitudes (Figure 7b,e,h). In both models, zonal wind is reduced in the exit region but it increases somewhat on the equatorward and poleward sides of the same region. In CESM, the refinement-induced changes are weak and do not generally improve the biases (Figure 7a,d,g). Over the North Pacific, changes are mostly insignificant except in MPAS, where the jet weakens on the poleward side of the entrance region (Figure 7g–i), improving the bias there (Figure 7c,f). Though MPAS maintains the highest area-mean RMSE, refinement results in more comparable performance to the other models (Figure 7d–f).

The zonal-mean NH jet is too weak on the equatorward side of the climatological jet maximum and too strong on the poleward side in the unrefined runs (Figure 8a–c), in line with Figure 7a–c. This is seen at all levels, but the biases are strongest in the upper troposphere and lower stratosphere, consistent with the cold bias in the polar stratosphere (Figure 3). With refinement, the jet weakens across all levels in MPAS, with the strongest

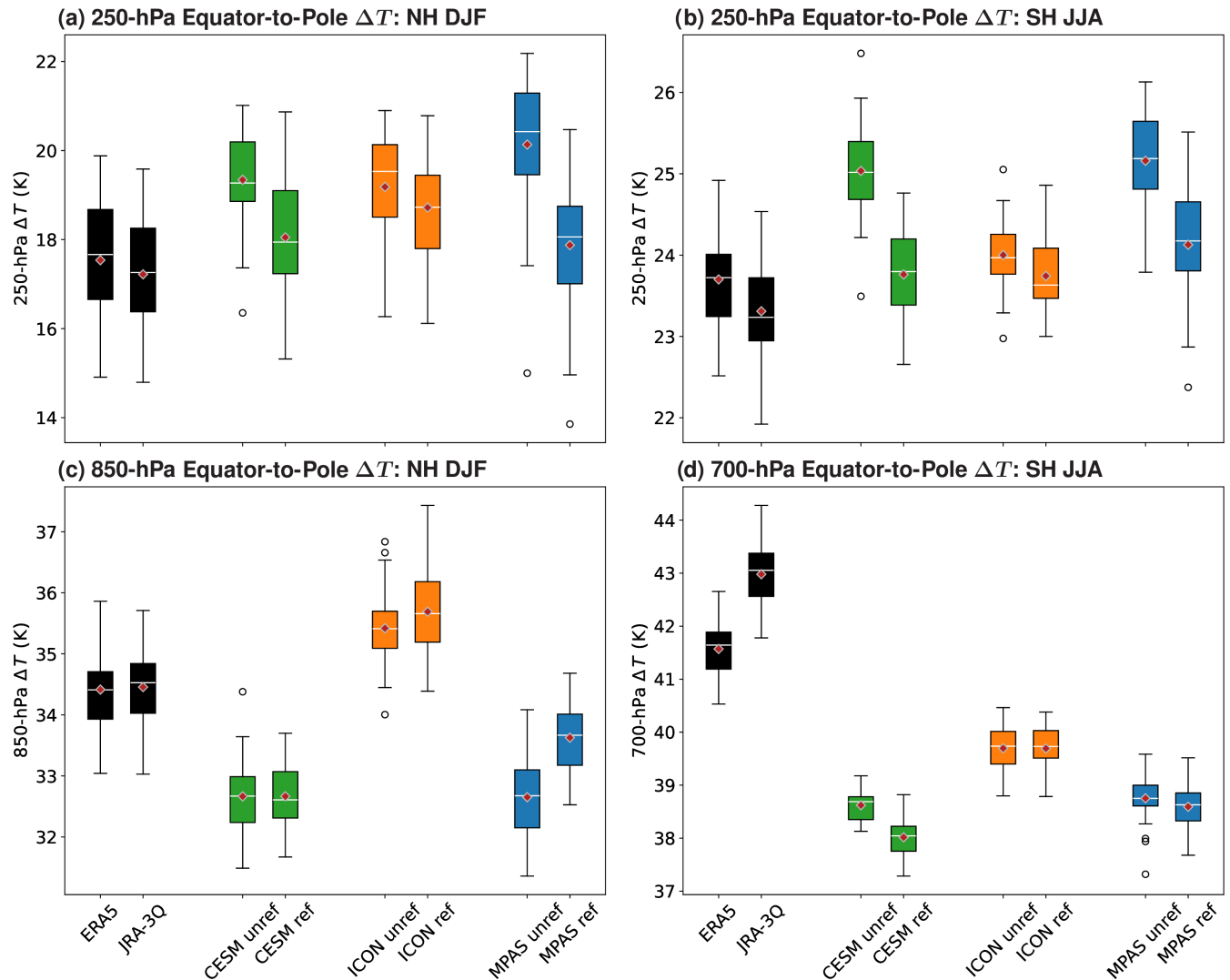


FIGURE 6 The (a, c) Northern Hemisphere (NH) and (b, d) Southern Hemisphere (SH) winter (December–February [DJF]) and June–August [JJA]) Equator-to-Pole temperature contrasts at (a, b) upper levels (250 hPa) and (c, d) lower levels (850 hPa in the NH and 700 hPa in the SH) for the fifth generation of the European Center for Medium-Range Forecasts Reanalysis (ERA5) and Japanese Reanalysis for Three Quarters of a Century (JRA-3Q; black boxes) and for the unrefined and refined runs for the Community Earth System Model (CESM; green boxes), the Icosahedral Nonhydrostatic (ICON) model (orange boxes), and the Model for Prediction Across Scales (MPAS; blue boxes). The contrasts are computed as the difference between the area-mean temperature in the Tropics (15° S– 15° N) and for the respective polar cap ($>60^{\circ}$ N or $>60^{\circ}$ S). The box and whiskers show the distribution of the individual seasonal means. The boxes extend from the first to the third quartiles, with the median shown as the white horizontal line and the mean as the pink diamond. The whiskers extend to the farthest data point or 1.5 times the interquartile range from the box edge; in the latter case, outliers are shown as open circles. Units are kelvin.

changes in the stratosphere (Figure 8i). This happens in a region where the jet is too strong in the unrefined runs, thus substantially improving its representation at all levels (Figure 8c,f,i). We find similar results for ICON, though the changes with the refinement are weaker and insignificant (Figure 8b,e,h), consistent with the 850 hPa zonal wind results (Figure 7b,e,h). Refinement also weakens the jet on the poleward side in CESM, but here this corresponds to a region where the jet is well represented or underestimated in the unrefined run, resulting in a worsening of

the negative bias at and slightly north of the jet center (Figure 8a,d,g).

In the SH, the 850 hPa jet is too strong in the unrefined runs (Figure 9a–c), similar to in the NH (Figure 7a–c), but with more model-dependent bias patterns. The overall performance is best in ICON (Figure 9b), where the most pronounced regional bias is an overestimation of the jet south and southeast of Australia, a small negative bias over the eastern South Pacific and South Atlantic, and a small positive bias over the Indian Ocean. MPAS exhibits

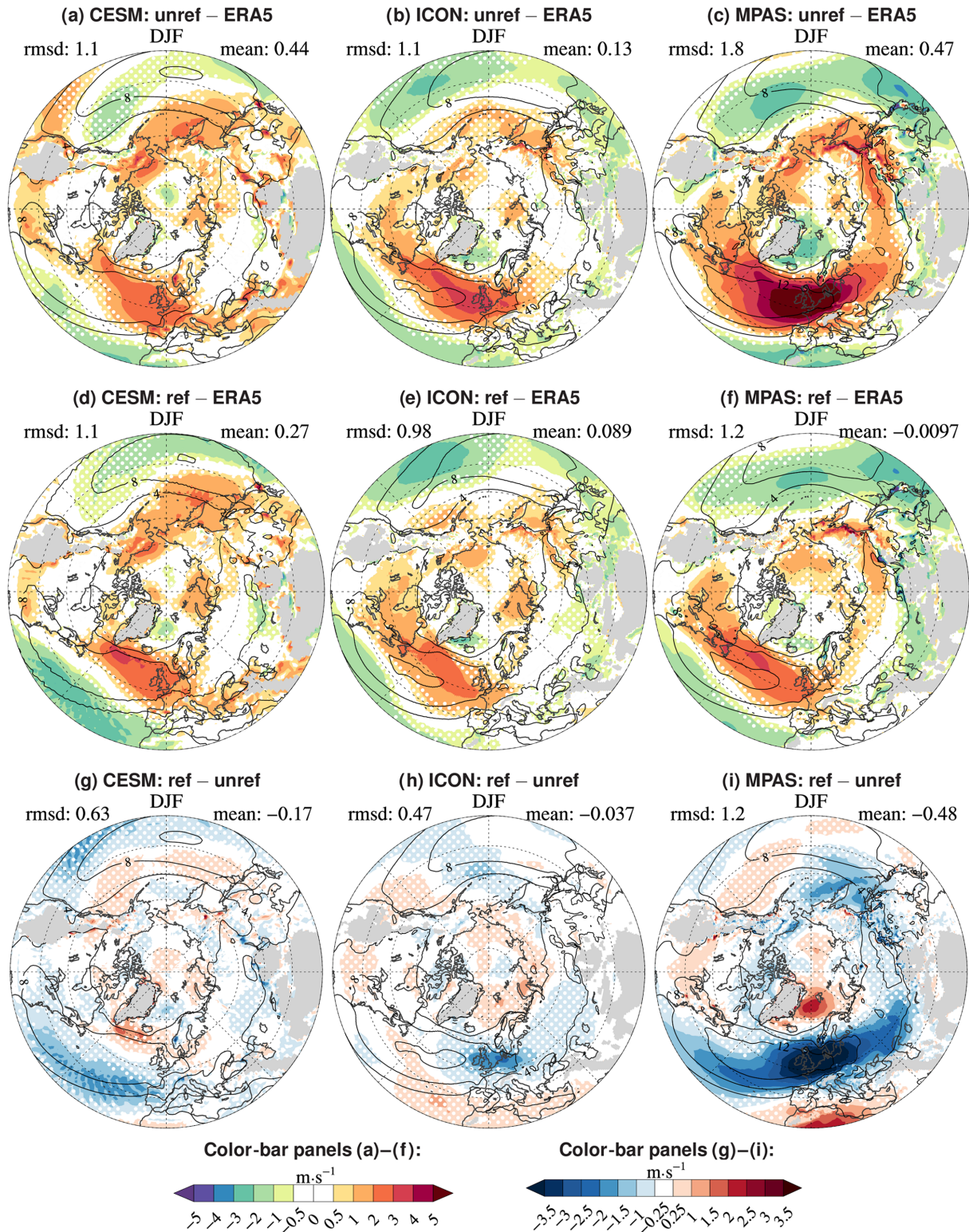


FIGURE 7 As in Figure 2, but Northern Hemisphere winter (December–February, DJF) 850 hPa zonal wind. Units are m.s^{-1} .

the largest errors, characterized by too strong winds in the Atlantic and Indian Ocean sectors and close to Antarctica and too weak winds over the South Atlantic on the equatorward side and in a small region downstream of New

Zealand (Figure 9c). The positive bias around coastal East Antarctica in MPAS actually means that its representation of the climatological coastal easterlies is too weak. In CESM, the jet is too strong on the poleward side and too

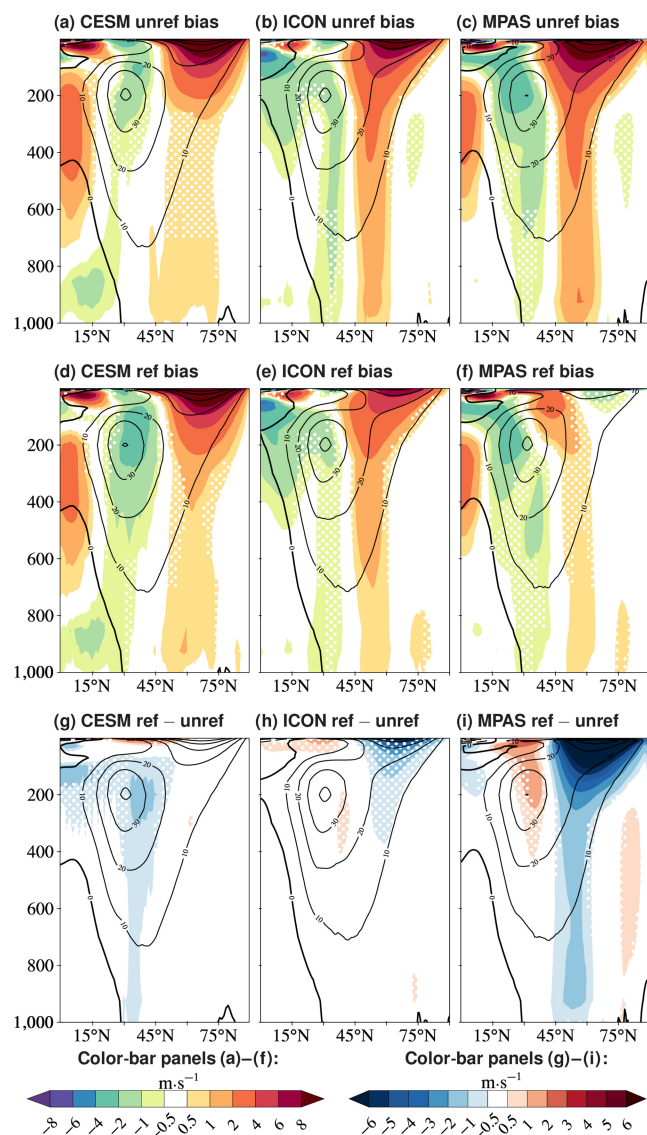


FIGURE 8 As in Figure 3, but for Northern Hemisphere winter (December–February) zonal-mean zonal wind. Units are $\text{m}\cdot\text{s}^{-1}$.

weak on the equatorward side in the Atlantic and Indian Ocean sectors (Figure 9a).

Refinement improves the SH 850 hPa jet in all models (Figure 9). Again, the strongest changes occur in MPAS, where the jet weakens on the poleward side across all longitudes, with the most pronounced weakening over the Indian Ocean (Figure 9g–i). This improves the biases over the South Atlantic, the 30°–60° S latitude band over the Indian Ocean, and adjacent to the Antarctic continent, where the jet was too strong in the unrefined MPAS run (Figure 9c,f). The improved Antarctic coastal easterlies in the refined MPAS run might simply indicate reduced sub-grid orography effects. Performance, however, deteriorates over the southern Indian Ocean and over the southeastern Pacific, where biases were small in the unrefined run

(Figure 9c,f,i); the overall effect is still improved area-mean RMSE and bias. In ICON, refinement slightly intensifies the jet over the South Atlantic while weakening it over the Weddell Sea and downstream across the southeastern Indian Ocean and over towards the southsoutheast of Australia (Figure 9h), improving the biases in the same regions (Figure 9b,e,h). Though changes in CESM are mostly weak and insignificant, the slight increase on the equatorward side and decrease on the poleward side in the Atlantic sector ameliorates the biases in this region and the area-mean bias and RMSE (Figure 9a,d,g).

The zonal-mean SH jet is too strong on the poleward side in all models in the unrefined runs, particularly in MPAS, and underestimated on the equatorward side in MPAS and ICON (Figure 10a–c). ICON performs best overall, especially in the mid and low troposphere. With refinement, there is a pronounced weakening on the poleward side in MPAS, which is relatively barotropic throughout the troposphere and amplified in the lower stratosphere (Figure 10i). The SH jet also weakens somewhat on the poleward side in ICON and CESM (Figure 10g–h), counteracting the positive biases in the unrefined runs (Figures 10a,d,g and 10b,e,h). In ICON, we additionally find intensification on the equatorward side in the mid and upper troposphere that improves a negative bias in the same region in the unrefined run (Figure 10b,e,h). Overall, the refinement improves the zonal-mean SH jet in all models (Figure 10); however, in MPAS, the zonal winds weaken so much on the poleward side that the bias changes sign (from positive in the unrefined run to negative in the refined run; Figure 10c,f,i), consistent with the results at 850 hPa (Figure 9c,f,i).

During summer, the North Atlantic and North Pacific 850 hPa jets are placed too far north in the unrefined runs with consistent biases in the zonal-mean NH jet (Supporting Information Figure S9a–c and S10a–c). In the SH, both the regional 850 hPa zonal wind biases and the sign of the area-mean bias vary across models (Supporting Information Figure S11a–c), as do the biases in the zonal-mean winds (Supporting Information Figure S12a–c). The refinement-induced changes yield some regional improvements, like an improved positive bias on the poleward side of the North Atlantic jet in MPAS (Supporting Information Figure S9c,f) and an improved negative bias on the equatorward side of the North Pacific jet in CESM (Supporting Information Figure S9a,d), but also degrades some features, like the negative bias on the equatorward side of the SH jet in CESM (Supporting Information Figure S12a,d).

In SH winter and summer, changes in zonal wind in MPAS and ICON (Figures 9 and 10, Supporting Information Figures S11 and S12) appear consistent with the changes in the 250 hPa temperature contrasts (Figure 6

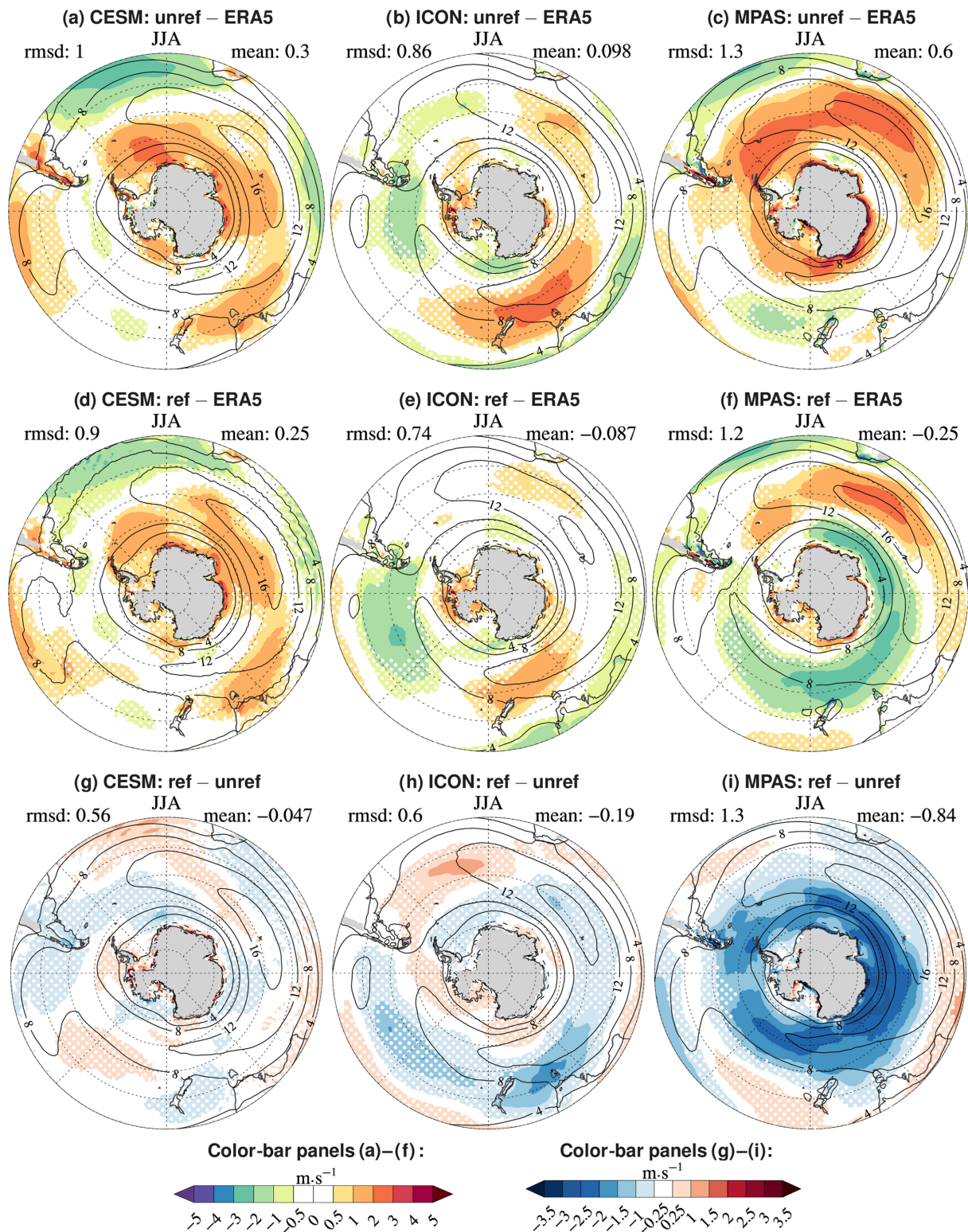


FIGURE 9 As in Figure 2, but for Southern Hemisphere winter (June–August, JJA) 850 hPa zonal wind. Units are m.s^{-1} .

and Supporting Information Figure S8), with strong reductions in both the contrasts and zonal wind in MPAS and similar but weaker changes in ICON. In CESM, the zonal-wind changes are weaker and limited to upper

levels (above 400 hPa; Figure 10), in spite of the pronounced changes in the 250 and 700 hPa contrasts. In NH winter, changes in zonal wind appear driven by local changes in baroclinicity, consistent with Harvey

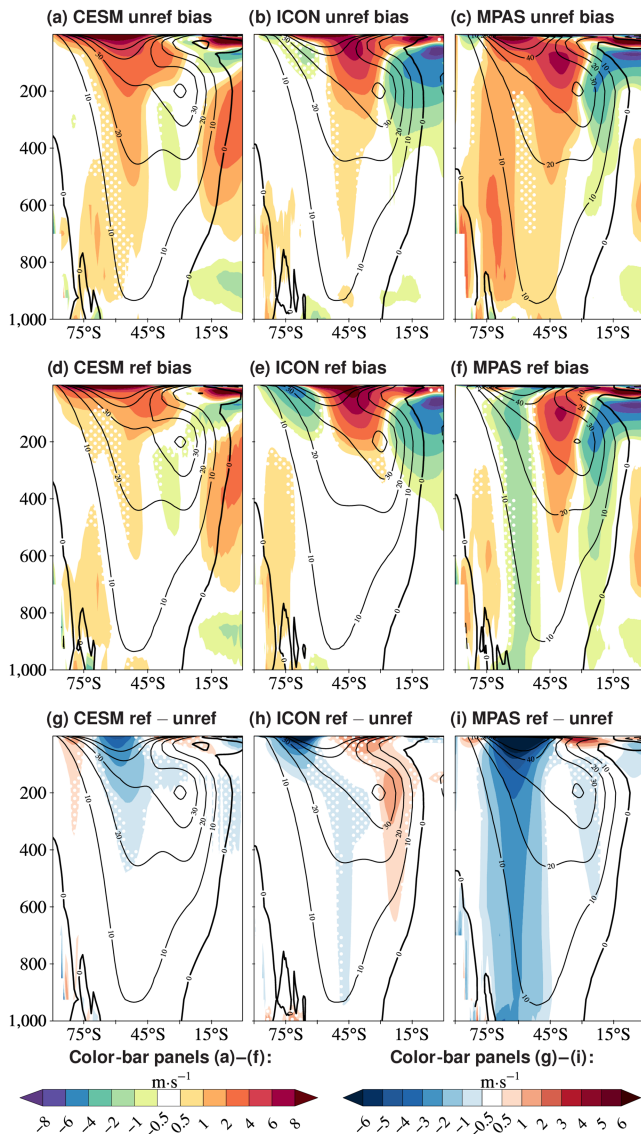


FIGURE 10 As in Figure 3, but for Southern Hemisphere winter (June–August) zonal-mean zonal wind. Units are m s^{-1} .

et al. (2014), who found the North Atlantic storm-track changes to be more sensitive to the local temperature gradient. Indeed, the 850 hPa zonal wind changes (Figure 7) align with the changes in the maximum Eady growth parameter (Eady, 1949; Lindzen & Farrell, 1980) over the North Atlantic basin (Supporting Information Figure S13). As for zonal wind, changes in the Eady parameter are most pronounced in MPAS, where Eady is too strong in the unrefined experiment and improves considerably with refinement. This is, furthermore, consistent with the refinement-induced changes in near-surface temperature, with warming over and east of the North American continent and cooling over and west of Europe (Figure 2i).

4.3 | Surface heat fluxes

The latent and sensible heat fluxes (hfls and hfss) relate to the near-surface wind speed (sfcWind), air and surface temperature ($ta(z)$ and ts), and specific humidity (hus) through the following bulk formulas (e.g., Foken, 2008):

$$hfss = C_H(z) \times sfcWind \times (ts - ta(z)), \quad (1)$$

$$hfls = C_E(z) \times sfcWind \times (hus(0) - hus(z)), \quad (2)$$

where z is height, C_H and C_E are transfer coefficients for heat and moisture, and the fluxes are positive from the surface to the atmosphere. Equations 1 and 2 imply that biases in hfls and hfss can be related to biases in surface and near-surface (2-m) temperature, near-surface (10-m) wind speed, and surface and near-surface (2-m) specific humidity. Biases in cloud cover and surface radiative fluxes and differences in boundary-layer parametrizations between models can also play a role, as related differences in wind velocity and stability impact the transfer coefficients. In the following, we consider the hfls and hfss biases in the context of these fields, focusing on the North Atlantic, North Pacific, NH high-latitude continents, and the SH ocean. The main bias characteristics are summarized in Supporting Information Tables S3 and S4.

We first examine the North Atlantic, focusing on the region north of the Gulf Stream and North Atlantic Drift (first part of Supporting Information Table S3). In ICON and MPAS, winter hfls and hfss are overestimated in this region, because the near-surface winds are too strong and the air is too cold and dry (Figures 12a–e and 13a–e). In MPAS, the cold bias coincides with negative rlds and cloud cover biases and positive rsds biases (Figure 13e–g; Supporting Information Table S3). We find similar, albeit weaker, biases in rlds and rsds in ICON; however, in this model, the cloud cover is overestimated, suggesting cloud property errors such as the clouds being too optically thin and transmissive (Figure 12f,g; Supporting Information Table S3). In CESM, the positive hfls biases stem from strong near-surface winds and too dry air (Figure 11a–c). CESM also overestimates cloudiness, with consistent positive biases in rlds and negative biases rsds; this, in combination with the warm near-surface bias noted in Section 4.1, contributes to an underestimation of hfss (Figure 11d–g; Supporting Information Table S3).

Over the North Pacific, hfls is underestimated in the climatological maximum north of the Kuroshio Current, and overestimated elsewhere, particularly in CESM (Figures 11a, 12a, and 13a). In what follows, we focus on the regions where hfls is too strong (second

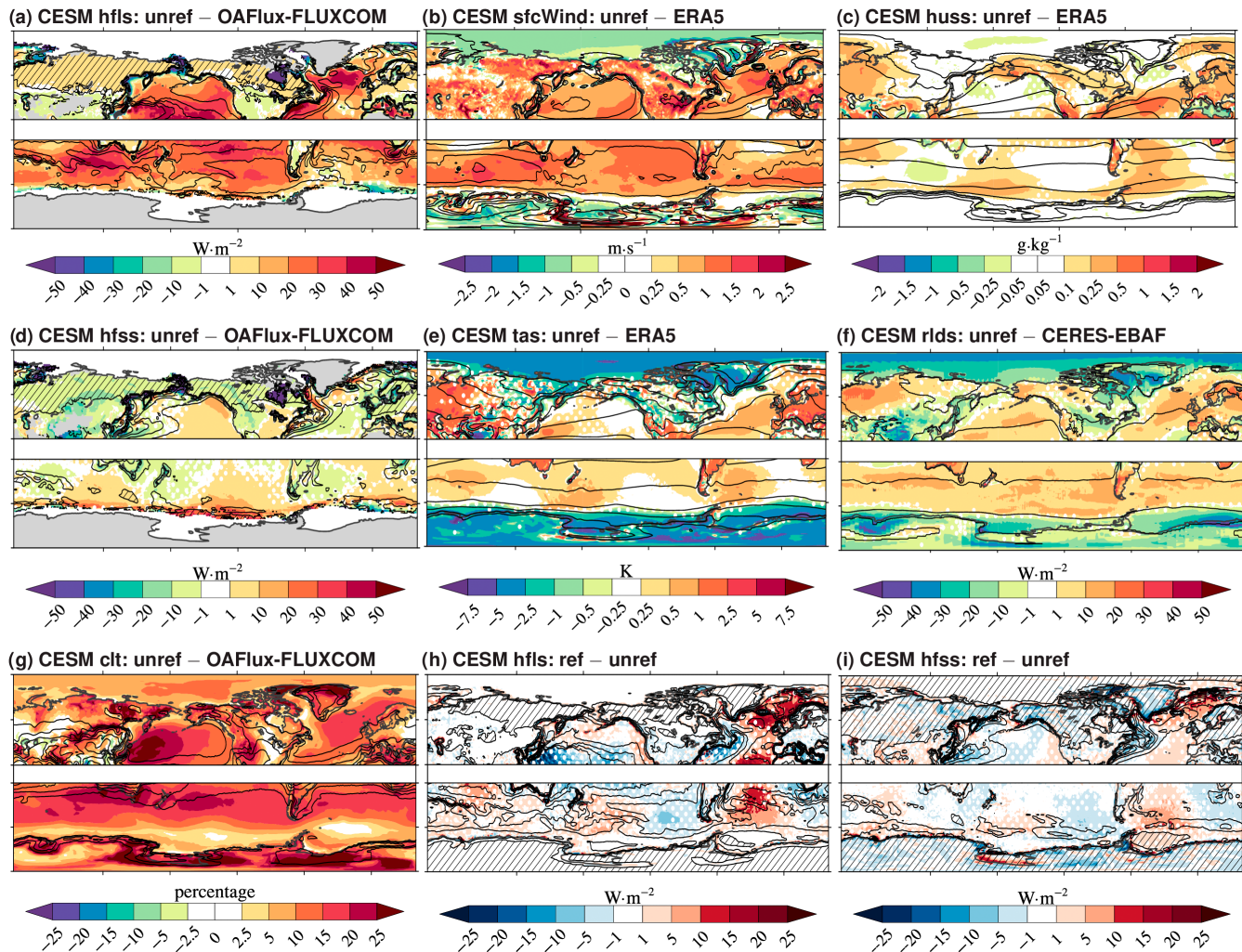


FIGURE 11 Winter biases in the unrefined the Community Earth System Model (CESM) runs for (a) latent heat flux (hfls; $\text{W}\cdot\text{m}^{-2}$), (b) near-surface wind speed (sfcWind; $\text{m}\cdot\text{s}^{-1}$), (c) near-surface specific humidity (huss; $\text{g}\cdot\text{kg}^{-1}$), (d) sensible heat flux (hfss; $\text{W}\cdot\text{m}^{-2}$), (e) near-surface temperature (tas; K), (f) surface long-wave radiative flux (rlds; $\text{W}\cdot\text{m}^{-2}$), and (g) cloud cover (clt; percent), shown with the differences between the refined and unrefined runs for (h) hfls and (i) hfss ($\text{W}\cdot\text{m}^{-2}$). In each panel, latitudes poleward of 30°N show the Northern Hemisphere (NH) winter (December–February) field and latitudes poleward of 30°S show the Southern Hemisphere (SH) winter (June–August) field. Biases are computed with respect to (a, d) OAFlux over ocean and FLUXCOM over land (1985–2014), (b, c, e) the fifth generation of the European Center for Medium-Range Forecasts Reanalysis (ERA5; 1985–2014), and (f) Clouds and the Earth's Radiant Energy System Energy Balanced and Filled (CERES-EBAF; 2001–2024). Black contours show the reference data climatology in (a) and (d) and the respective unrefined climatology elsewhere; regions where the climatological values are negative are indicated by hatching. White dots are as in Figure 2.

part of Supporting Information Table S3). In CESM, the positive hfls biases likely reflect a combination of overestimated near-surface winds and an excessively large C_E (Figure 11a,b). The hfss bias is only slightly positive because the near-surface temperatures are too warm (consistent with too strong rlds; Supporting Information Table S3), reducing the difference between surface and near-surface levels, given that the SST is prescribed (Equation (1); Figure 11d–f). In ICON and MPAS, the more moderate hfls and hfss biases could be associated with overly large C_E and C_H values due to issues with boundary-

layer parametrizations, given the small biases in wind, humidity, and temperature (Figures 12a–e and 13a–e).

The hfls and hfss biases over NH land are generally smaller compared with over ocean (Figures 11a,d, 12a,d, and 13a,d). All models overestimate hfls over high-latitude land and underestimate hfss (Figures 11a,d, 12a,d, 13a,d, third part of Supporting Information Table S3). As the FLUXCOM climatology is negative (hatching in Figures 11–13) over high-latitude land during winter, the negative hfss biases are consistent with an overestimation of the flux from the atmosphere to the surface

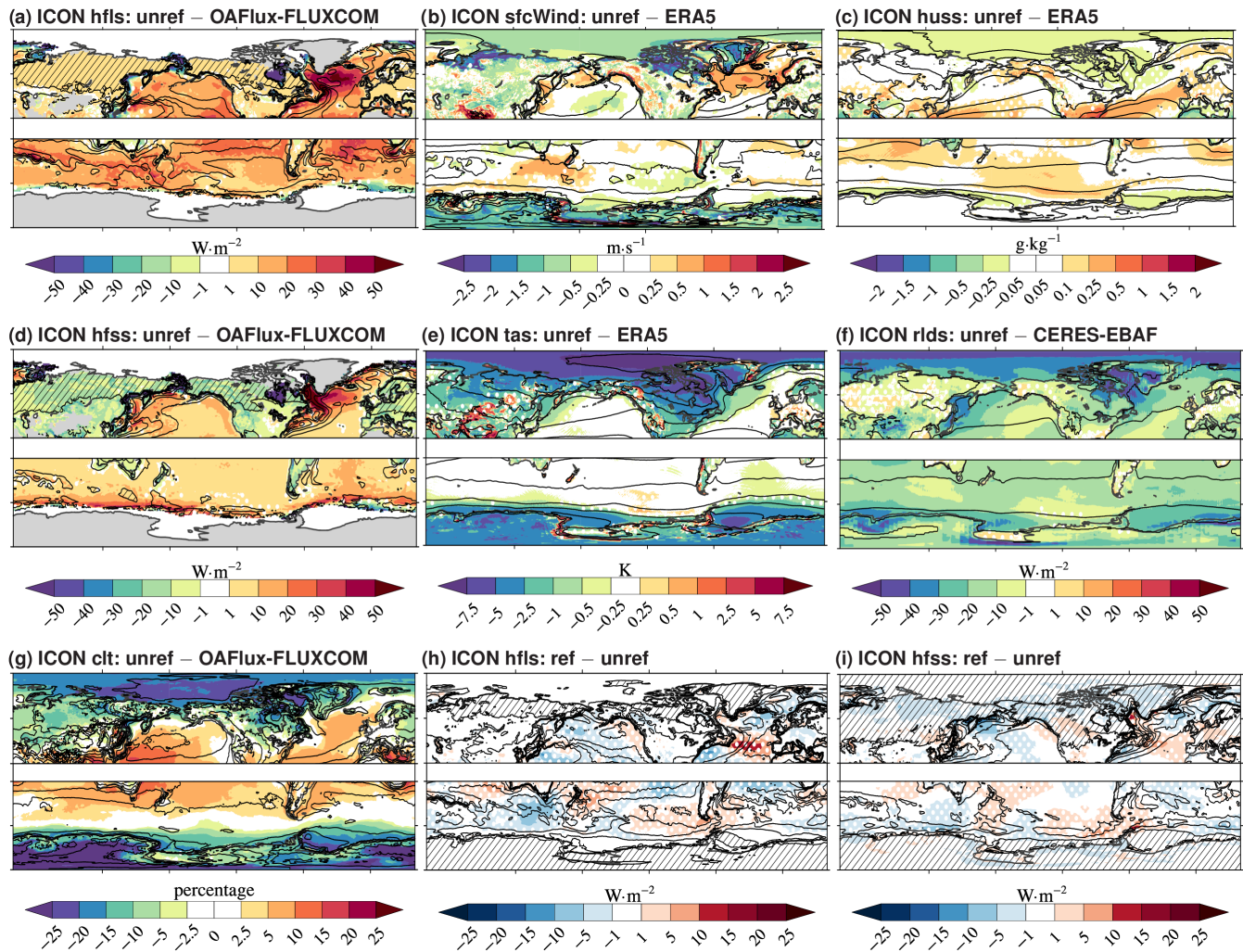


FIGURE 12 As in Figure 11, but for the Icosahedral Nonhydrostatic (ICON) model for winter (December–February in the Northern Hemisphere and June–August in the Southern Hemisphere). In (h) and (i), the Arctic refined run is used in the NH and the Antarctic refined run in the SH.

(Figures 11d, 12d, and 13d). This is consistent with the negative near-surface temperature biases in the same regions in ICON and MPAS, related to underestimated cloudiness and rlds (Figures 12d–g and 13d–g). The positive hfls biases, on the other hand, are consistent with the flux from the atmosphere to the surface being too weak (Figures 11a, 12a, and 13a). The lack of consistent biases in hfls, humidity, and wind indicate that the surface could be releasing too much latent energy as a result of too active sublimation (Figures 11a–c, 12a–c, and 13a–c).

The most pronounced and significant differences between the refined and unrefined experiments in the NH are generally found for hfls over the North Atlantic (Figures 11h,i, 12h,i, and 13h,i). In MPAS, both hfls and hfss decrease over the northwestern part of the North Atlantic and increase over the southeastern part, similar to the bias patterns in near-surface temperature and specific humidity (Figure 13c,e,h,i). Refinement-induced changes in the area-mean RMSE values for the NH extratropics

are summarized in Figure 14a (further discussed in Section 4.4). The refinement slightly improves the representation of hfls over ocean and land in MPAS, and over land in ICON (Figure 14a). For hfss, the refinement yields improved area-mean RMSE values over ocean in MPAS (Figure 14a).

Over the SH ocean (fourth part of Supporting Information Table S3), we predominantly find positive winter hfls biases, with the largest values in CESM (Figures 11a, 12a, and 13a). This might be related to exaggerated turbulent mixing due to strong winds, which again are strongest in CESM (Figures 11b, 12b, and 13b) and to an overestimated C_E . Around the Antarctic sea-ice edge, we find positive and negative biases in all models (Figures 11a, 12a, and 13a), potentially reflecting differences in sea-ice extent between the models and OAF flux. We also find positive hfss biases in MPAS and ICON, similar to in the NH (Figures 12d and 13d). In MPAS, positive near-surface wind speed biases and an overestimated C_H

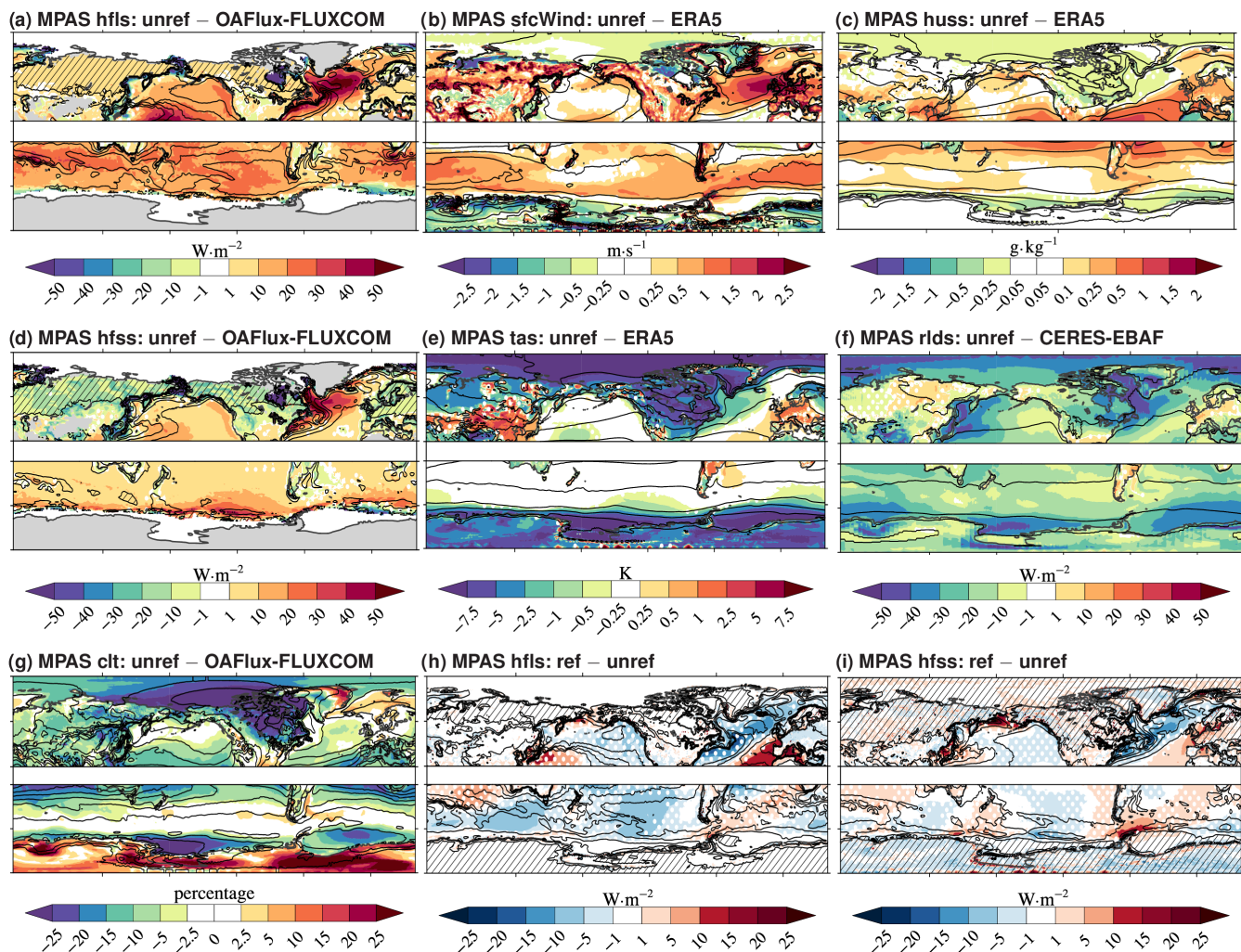


FIGURE 13 As in Figure 11, but for the Model for Prediction Across Scales (MPAS) for winter (December–February in the Northern Hemisphere and June–August in the Southern Hemisphere). In (h) and (i), the Arctic refined run is used in the NH and the Antarctic refined run in the SH.

could contribute to the stronger turbulent mixing and enhanced hfss (Figure 13b,d). In ICON, both wind speed and temperature biases are small, suggesting a considerable contribution from C_H (Figure 12b,e). CESM underestimates hfss where near-surface temperature is too high (Figure 11d,e). As noted in Section 4.1, the warm bias in near-surface temperature in CESM is accompanied by overestimated cloudiness and rlds (Figure 11e–g). In general, hfss biases are smaller than the hfls biases, except near the sea-ice edge (again likely due to differences between the sea-ice extent in the models and OAFflux).

The strongest refinement-induced changes in the SH is a reduction in hfls over the southern Pacific and southern Indian Ocean in MPAS and an increase over the South Atlantic in CESM, improving regional biases in the former case (Figures 11a,h,i and 13a,h,i). Overall, the area-mean RMSE values for the SH extratropics (summarized in Figure 15a) improve with refinement for hfls in MPAS and ICON, whereas performance in CESM degrades somewhat

(Figure 15a). For hfss, changes are weaker and generally not associated with improved area-mean RMSE values (Figures 11i, 12i, and 13i; Figure 15a).

During summer (Supporting Information Table S4), MPAS and ICON overestimate hfls over most of the NH and SH oceans, except downstream of NH high-latitude continents, where the (negative) fluxes are too weak (Supporting Information Figures S15a and S16a), potentially tied to too strong C_E and, in MPAS, to too strong winds over the SH ocean (Supporting Information Figure S16b). CESM underestimates hfls in the northwestern Atlantic, North Pacific, and southwestern Atlantic, likely linked to positive specific humidity biases in these regions (Supporting Information Figure S14a,c). Furthermore, CESM overestimates hfls over all other ocean regions, likely linked to too strong C_E and too strong winds over the SH ocean (Supporting Information Figure S14a–b). Over the ocean, hfss biases are positive everywhere, entailing that the fluxes are too weak at

(a) NH extratropics DJF

NH ET		CESM	CESM	ICON	ICON	MPAS	MPAS
		unref	ref-unref	unref	ref-unref	unref	ref-unref
ta250	DJF	1.8	-0.41	1.8	-0.20	2.3	-1.1
ta500	DJF	0.81	0.010	1.1	0.036	1.1	-0.30
ta850	DJF	1.1	0.00	1.3	0.30	1.4	-0.42
tas	DJF	1.9	-0.12	2.5	0.15	3.3	0.39
ua250	DJF	2.3	0.36	2.2	-0.13	3.5	-1.4
ua850	DJF	1.1	0.065	1.1	-0.093	1.8	-0.62
sfcWind	DJF	1.0	-0.17	0.77	-0.0056	1.3	-0.23
huss	DJF	0.34	-0.042	0.22	0.0058	0.35	-0.021
hfls-ocn	DJF	33	0.68	28	-0.0063	28	-1.8
hfls-lnd	DJF	5.4	0.090	5.2	-0.16	7.0	-0.24
hfss-ocn	DJF	34	0.22	34	0.38	32	-1.1
hfss-lnd	DJF	9.8	0.028	12	0.26	13	-0.046
psl	DJF	2.4	-0.086	2.1	0.032	4.9	-1.5
Counts	Best/ Better	4	5	8	5	1	11
Counts	Worst/ Worse	2	6	0	7	11	1

(b) Arctic DJF

Arctic		CESM	CESM	ICON	ICON	MPAS	MPAS
		unref	ref-unref	unref	ref-unref	unref	ref-unref
ta250	DJF	3.2	-0.82	3.1	-0.44	4.0	-2.5
ta500	DJF	0.81	-0.058	1.3	0.013	1.2	-0.40
ta850	DJF	1.2	-0.17	1.6	0.47	1.5	-0.51
tas	DJF	2.5	-0.11	4.1	0.083	4.9	0.83
ua250	DJF	1.8	0.20	1.3	0.043	2.2	-1.0
ua850	DJF	0.97	0.071	0.70	0.052	1.3	-0.44
sfcWind	DJF	1.0	-0.20	1.1	-0.030	1.5	-0.27
huss	DJF	0.20	-0.060	0.11	0.00	0.15	0.011
hfls-ocn	DJF	35	2.4	34	-0.13	32	-1.2
hfls-lnd	DJF	5.1	-1.0	4.4	-0.18	6.6	0.21
hfss-ocn	DJF	71	-0.040	67	1.0	64	-0.32
hfss-lnd	DJF	8.4	1.2	13	1.4	15	0.56
psl	DJF	3.2	-0.19	2.5	-0.079	7.9	-2.8
Counts	Best/ Better	5	8	6	4	2	8
Counts	Worst/ Worse	3	4	2	8	8	4

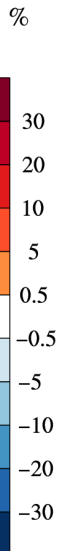


FIGURE 14 Overview of (a) Northern Hemisphere (NH) extratropical and (b) Arctic winter (December–February, DJF) root-mean-square error (RMSE) values for 250 hPa, 500 hPa, 850 hPa, and near-surface temperature (ta250, ta500, ta850, and tas; K), 250 hPa and 850 hPa zonal wind (ua250 and ua850; $\text{m}\cdot\text{s}^{-1}$), near-surface wind speed (sfcWind; $\text{m}\cdot\text{s}^{-1}$), near-surface specific humidity (huss; $\text{g}\cdot\text{kg}^{-1}$), surface latent and sensible heat flux (hfls and hfss; $\text{W}\cdot\text{m}^{-2}$), and sea-level pressure (psl; hPa). Each panel shows the variable (column 1), season (column 2), and RMSE values for the unrefined runs (columns titled “unref”) and the difference between the refined and unrefined (columns titled “ref – unref”) for the Community Earth System Model (CESM; columns 3 and 4), the Icosahedral Nonhydrostatic (ICON) model (columns 5 and 6), and the Model for Prediction Across Scales (MPAS; columns 7 and 8). For each variable (row), the smallest (largest) unref value is underlined (shown in bold). The “ref – unref” values are highlighted by color shading indicating the percentage change relative to the unrefined value (color bar). The bottom two rows summarize the number of underlined (best) and bold (worst) unref values, and the number of ref – unref values that improve by more than -0.5% (blue cells; better) and deteriorate by more than 0.5% (red cells; worse). Reference data are the European Center for Medium-Range Forecasts Reanalysis (ta250, ta500, ta850, tas, ua250, ua850, sfcWind, huss, psl), OaFlux (hfls-ocn and hfss-ocn), and FLUXCOM (hfls-lnd and hfss-lnd).

high latitudes (where climatological fluxes are negative) and too strong at lower latitudes (where climatology is positive) in both hemispheres (Supporting Information Figures S14d, S15d, and S16d). In CESM and MPAS, the low-latitude biases in the SH are likely due to strong winds and overestimated C_H . Continental NH hfss biases are more pronounced, with fluxes being too weak at high latitudes and too strong further south (Supporting Information Figures S14d, S15d, and S16d), potentially reflecting uncertainties in C_H or biases in the temperature differences ($t_s - t_a$).

We find significant changes of both signs with refinement for hfls during summer (Supporting Information Figures S14h, S15h, and S16h) that improve the area-mean

RMSE over the oceans for MPAS and ICON (Supporting Information Figures S17a and S18a). For hfss, changes are strongest over high-latitude NH land (Supporting Information Figures S14i, S15i, and S16i), yielding improved area-mean RMSE over NH land in ICON and CESM (Supporting Information Figure S17a).

4.4 | Model performance overview

Here, we synthesize the benefits of the refinement in terms of area-mean RMSE for temperature and zonal wind at various levels, near-surface wind speed and specific humidity, hfls, hfss, and sea-level pressure (Figures 14

(a) SH extratropics JJA

SH ET		CESM	CESM	ICON	ICON	MPAS	MPAS
		unref	ref-unref	unref	ref-unref	unref	ref-unref
ta250	JJA	1.6	-0.56	1.1	-0.0064	1.7	-0.33
ta500	JJA	1.0	-0.28	1.0	-0.068	1.1	-0.23
ta850	JJA	1.4	-0.12	0.91	-0.048	0.62	-0.077
tas	JJA	1.6	-0.099	1.5	0.088	2.7	-0.31
ua250	JJA	2.3	-0.50	1.8	-0.39	2.9	0.17
ua850	JJA	1.0	-0.13	0.86	-0.12	1.3	-0.12
sfcWind	JJA	1.0	-0.14	0.55	-0.021	0.96	-0.20
huss	JJA	0.21	-0.025	0.15	0.0055	0.30	0.046
hfls-ocn	JJA	24	0.70	18	-0.43	19	-1.1
hfss-ocn	JJA	13	0.34	13	-0.054	13	0.33
psl	JJA	2.9	-0.30	4.2	-0.28	8.7	-2.5
Counts	Best/ Better	2	9	7	8	2	8
Counts	Worst/ Worse	3	2	1	2	7	3

(b) Antarctic JJA

Antarctic		CESM	CESM	ICON	ICON	MPAS	MPAS
		unref	ref-unref	unref	ref-unref	unref	ref-unref
ta250	JJA	2.3	-0.97	1.4	-0.31	2.1	-0.98
ta500	JJA	1.3	-0.48	1.2	-0.094	1.4	-0.46
ta700	JJA	1.5	-0.30	1.5	-0.052	1.1	-0.19
tas	JJA	2.8	-0.19	2.8	0.17	5.2	-0.60
ua250	JJA	1.5	-0.24	0.97	-0.031	1.7	0.065
ua700	JJA	1.4	-0.27	0.96	-0.063	2.1	-0.77
sfcWind	JJA	1.00	-0.20	0.90	-0.017	1.3	-0.30
huss	JJA	0.17	-0.028	0.11	0.00	0.14	0.00
hfls-ocn	JJA	18	0.16	16	0.040	17	-0.73
hfss-ocn	JJA	33	0.50	29	0.077	30	0.48
psl	JJA	4.4	-0.26	7.8	-0.49	16	-5.0
Counts	Best/ Better	1	9	9	8	1	9
Counts	Worst/ Worse	5	2	0	1	6	2

%

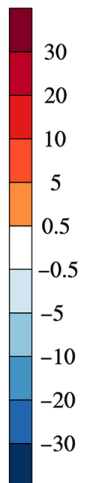


FIGURE 15 As in Figure 14, but for the (a) Northern Hemisphere (SH) extratropics (a) and (b) the Antarctic during winter (June–August, JJA). Note that we use the 700 hPa surface for ta and ua in (b), to ensure that we are above the Antarctic topography and that we omit comparison with FLUXCOM (hfls-F and hfss-F in Figure 14).

and 15 and Supporting Information Figures S17–S22), focusing on variables for which the errors improve the most. We use RMSE because it emphasizes large errors and prevents cancellation between positive and negative values. To assess whether the changes are more or less pronounced when only considering the region with the highest resolution, we include values for the Arctic and Antarctic ($>60^\circ$ N/S) alongside those for the NH and SH extratropics ($>30^\circ$ N/S).

MPAS benefits the most from the refinement in the NH during winter (Figure 14). We find temperature (at 250, 500, and 850 hPa), zonal wind (at 250 and 850 hPa), and sea-level pressure to improve the most, both in terms of absolute values (numbers) and relative values (color shading), and both for the NH extratropics and the Arctic (Figure 14). The most extreme improvement occurs for 250 hPa temperature, which improves by 48% in the NH extratropics and 63% in the Arctic. MPAS also improves its representation of 250 hPa and 850 hPa zonal wind by 40% and 34% respectively (Figure 14a), with the largest changes occurring in the North Atlantic region, where baroclinicity improves considerably (Figure 7c,f,i and Supporting Information Figure S13c,f,i). The heat fluxes also improve the

most over the North Atlantic Ocean (Figure 13a,d,h,i), but the area-mean RMSE changes are generally more modest (Figure 14a). For SH winter (Figure 15), MPAS generally improves for the same variables, except 250 hPa zonal wind and hfss over ocean. However, the magnitude of changes are more comparable to the other models in the SH, particularly for the extratropics.

CESM improves for temperature, wind, specific humidity, and sea-level pressure in the SH during winter in both the extratropics and Antarctic (Figure 15). The most pronounced improvements occur in high-latitude 250 and 500 hPa temperature, with the Antarctic 250 hPa temperature improving by 42%. Consistent with improved meridional temperature contrasts (Figure 6), the extratropical 250 hPa zonal wind improves by 22%. Improvements are more sparse in the NH during winter; however, the amelioration of the high-latitude upper-level temperature errors (cold bias) is clearly seen also here, with Arctic 250 hPa temperature improving by 26%; furthermore, Arctic continental hfls improves by nearly 20% (Figure 14), possibly associated with improved representation of near-surface wind and specific humidity.

Improvements for ICON are largely weaker. In the SH during winter, ICON improves to some extent for all fields except near-surface temperature and specific humidity (Figure 15). The strongest changes are found for Antarctic 250 hPa temperature, which improves by 22% (Figure 15b), and for 250 hPa and 850 hPa wind in the SH extratropics, which improve by 22% and 14% respectively (Figure 15a). Similar improvements in upper-level temperature and zonal winds are also evident for NH winter, albeit slightly weaker (Arctic 250-hPa temperature improving by 14% and NH extratropical 850 hPa zonal wind by 8%; Figure 14).

Overall, the most consistent improvement across models and hemispheres is the amelioration of the upper-level cold bias, with area-mean RMSE for 250 hPa temperature consistently improving in both hemispheres (Figures 14 and 15). MPAS has the largest changes in the NH and the Antarctic, whereas the models are more comparable in the SH extratropics. The number of variables that improve per model is similar in the SH, whereas in the NH extratropics MPAS tends to improve for more variables (second last row in Figures 14 and 15). MPAS also tends to have the largest errors in the unrefined run (bold values and last row in Figures 14 and 15). ICON, on the other hand, which has weaker improvements in many cases, tends to have smaller errors in the unrefined run (underlined values and second last row in Figures 14 and 15).

Results are more mixed during summer (Supporting Information Figures S17 and S18). Refinement yields the clearest benefits in the polar regions, both in terms of the magnitude of the RMSE reductions and the number of variables that improve, particularly for MPAS in the Arctic. As during winter, we find the most consistent improvements in 250 hPa temperature, with reduced RMSE in all regions and models except MPAS for the NH extratropics.

To assess the sensitivity of our results to the reference data, we produced a complementary set of RMSE plots based on JRA-3Q for temperature, zonal wind, near-surface wind, and sea-level pressure and ERA5 for hfls and hfss for winter (Supporting Information Figures S19 and S20) and summer (Supporting Information Figures S21 and S22). The results show that temperature, zonal wind, and sea-level pressure are largely insensitive to the choice of reference data, whereas other fields can be more sensitive. For near-surface wind, JRA-3Q has weaker values than ERA5, resulting in larger unrefined RMSE values in the complementary RMSE plots (Supporting Information Figures S19–S22) compared with the main RMSE plots (Figures 14 and 15 and Supporting Information Figures S17 and S18). This is possibly related to inappropriate terrain and inaccuracies in atmospheric boundary-layer processes in the data assimilation system

in JRA-3Q (Mi & Liu, 2025). For hfss and hfls, the unrefined biases computed with respect to OAFlux and FLUXCOM tend to be larger than those computed with respect to ERA5, for example over the NH ocean during winter.

4.5 | Broader polar impacts and areas for further investigation

Finally, we take a broader perspective on the influence of polar refinement by considering wider set of variables. Figure 16 highlights those most affected by regional refinement and that exhibit a high signal-to-noise ratio during winter. As we show polar-cap means, it should be kept in mind that regional variations within the polar cap, especially those with opposing signals, may result in a near-zero net signal, potentially obscuring significant localized impacts.

The intermodel differences are often larger than the differences between Arctic and Antarctic refinements within a single model. This suggests that the impact of the refinement is strongly model dependent, whereas it tends to show more uniformity when comparing the Arctic and Antarctic regions. Generally, MPAS changes the most and ICON the least, in line with findings from previous sections. Given the range of variables within Figure 16, not every variable can be discussed in detail. Instead, we focus on three key themes that emerged from the analysis: geopotential-height patterns, snow-depth changes, and cloud-cover variability. These themes illustrate distinct aspects of the refinement's impact and provide coherent narratives for further exploring the effects of increased resolution.

One of the most consistent effects of the refinement is the changes in geopotential height. The models tend to show an increase in polar geopotential height in both the troposphere (500 hPa) and stratosphere (50 hPa), indicating coupled changes across these layers. This effect is most pronounced in MPAS and is associated with a weakening of the zonal winds in the upper and lower troposphere, consistent with findings from Section 4.2. The question of whether the tropospheric jet weakening is driven by a weakened stratospheric polar vortex lies beyond the scope of this study but represents an interesting avenue for future research.

Another consistent effect of refinement is the change in snow depth. The models tend to simulate higher snow depth in both the Arctic and Antarctic with refinement. Unfortunately, MPAS snow-depth data were excluded due to processing errors; however, the intact parts of the MPAS output also suggest an increase in the refined version.

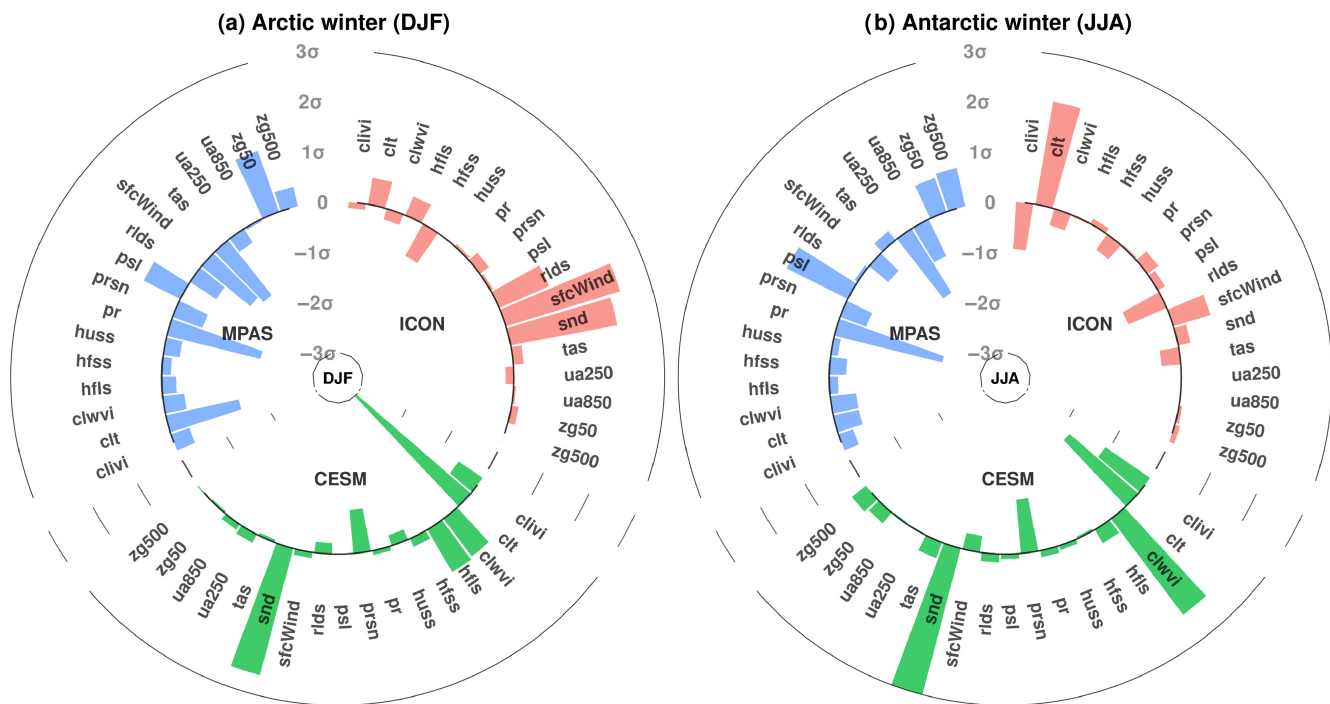


FIGURE 16 Overview of the response to polar refinement for the (a) Arctic and (b) Antarctic during winter (December–February [DJF] and June–August [JJA]). Each panel shows the normalized area-mean difference between the refined and unrefined experiments (colored bars) for the Model for Prediction Across Scales (MPAS; blue bars), the Community Earth System Model (CESM; red bars), and the Icosahedral Nonhydrostatic (ICON) model (green bars) for ice-water path (clwi), total cloud cover (clt), condensed-water path (clwvi), surface upward latent heat flux (hfls), surface upward sensible heat flux (hfss), near-surface specific humidity (huss), precipitation rate (pr), snowfall rate (prsn), sea-level pressure (psl), surface downwelling long-wave radiation (rlds), near-surface wind speed (sfcWind), snow depth (snd), near-surface temperature (tas), 250 hPa and 850 hPa zonal wind (ua250 and ua850), and 50 hPa and 500 hPa geopotential height (zg50 and zg500). The differences are normalized by the standard deviations of the interannual seasonal values from the relevant unrefined experiment, yielding standardized anomalies. The bars show the standardized anomalies in terms of standard deviations, with inward-oriented bars showing negative changes and outward-oriented bars showing positive changes. Units are percent.

The models diverge, however, with respect to total cloud cover. Whereas CESM and MPAS show strong decreases in polar cloud cover, ICON shows an increase, and these signals are consistent across hemispheres. Some evidence suggests that the models' base states may influence these differences. For example, CESM's overestimation of polar cloud cover results in reductions under refinement, producing more realistic values.

The three variables described tend to exhibit seasonal consistency. In the respective summer seasons (Supporting Information Figure S23), changes in snow depth remain largely unchanged across models. Resolution-induced changes in cloud cover are also largely consistent, with increased summer cloud cover in both hemispheres in ICON and decreased summer cloud cover in CESM and MPAS, except for a slight increase in CESM during Antarctic summer. Polar-cap mean geopotential heights also tend to increase in summer, although with reduced consistency and weaker magnitude compared with winter.

Overall, these findings underscore the complexity and model dependence of polar impacts resulting from regional refinement, highlighting areas where further investigation could yield valuable insights. In particular, the strong variability across models in cloud cover responses and the potential coupling between stratospheric and tropospheric changes warrant more detailed exploration. Additionally, the role of long-term snow accumulation versus seasonal snow processes in shaping snow-depth responses could benefit from targeted regional analyses. By delving deeper into these phenomena, future studies can clarify the mechanisms driving the diverse effects of refinement and their implications for polar climate processes.

5 | SUMMARY AND DISCUSSION

To explore the benefits and challenges of grid refinement in the polar regions, we conducted a coordinated

set of global variable-resolution model experiments for the recent past (1985–2014) with polar refinement (“the refined experiments”) and baseline experiments without refinement (“the unrefined experiments”) using three models: CESM, ICON, and MPAS. The refined experiments were configured with the finest horizontal resolution poleward of 60°, intermediate resolution in a transitional region equatorward of 60°, and coarsest resolution equatorward of the transitional region (Figure 1). The unrefined experiments were configured with the coarsest resolution everywhere. The result is the first coordinated multimodel variable-resolution experiment set with refinement over the polar regions. In this article, we document the experiment protocol and present the first results, focusing on the NH and SH extratropics and the polar regions for their respective winter seasons.

Upper tropospheric and lower stratospheric winter temperature consistently improves with refinement in the Arctic and Antarctic. The models are characterized by cold biases aloft in the polar regions in the unrefined runs and simulate warming in much the same regions with the refinement, thus ameliorating the errors there (Figures 3, 5, 14, and 15). MPAS exhibits the most dramatic improvement, with RMSE for Arctic 250 hPa temperature reduced by 63%. In line with improved upper-level cold biases, the upper-level (250 hPa) winter temperature contrasts between the Tropics and the Poles are also better represented in the refined runs (Figure 6).

The cold bias in the polar stratosphere and upper troposphere is a long-standing problem in climate modeling (Hamilton *et al.*, 1995). It is typically attributable to a too moist tropopause that causes overly strong radiative cooling (Bland *et al.*, 2021) in the upper troposphere–lower stratosphere region and a lack of dynamical variability and associated lack of dynamical warming at higher stratospheric levels (Charlton-Perez *et al.*, 2013; Lawrence *et al.*, 2022). The improved cold bias in the refined simulations is coupled to the weakened zonal winds via the thermal-wind balance. Enhanced gravity-wave drag associated with better-resolved orography is a plausible contributing mechanism wherein the stronger wave flux propagates upward and is dissipated in the upper atmosphere leading to weakened winds down to the lower levels and warmer polar upper atmosphere. However, we do not have the necessary output to diagnose gravity-wave tendencies directly and therefore cannot assess its quantitative role. Further experiments with detailed output from the gravity-wave parametrizations are needed to verify this hypothesis.

Consistent with improved large-scale and local-scale temperature contrasts, the winter zonal wind largely improves (Figures 7–10). The most striking improvement occurs in the North Atlantic region in MPAS, over and

downstream of the exit region where the jet is too strong in the unrefined run and considerably weaker and closer to ERA5 in the refined run. We also find similar improvements for ICON, albeit much weaker, but not for CESM. In MPAS and ICON, the North Atlantic improvements are linked to reduced biases in baroclinicity (Supporting Information Figure S13).

Near-surface winter cold biases generally persist with refinement in the Arctic and Antarctic compared with ERA5 (Figures 2 and 4). In the NH, near-surface temperature biases appear to be linked to biases in rlds and clouds (Supporting Information Figure S2), particularly in MPAS and ICON. Consistent biases in rlds and near-surface temperature are also seen in the Antarctic, albeit there the connection to clouds (Supporting Information Figure S3) is not so clear, indicating problems with turbulence parametrizations under stable stratification. The NH high-latitude near-surface temperature bias intensifies with refinement compared with ERA5 and is largely accompanied by consistent changes in rlds and clouds. In the SH, the influence of the refinement is more mixed, as biases improve in MPAS due to refinement-induced warming over Antarctica, and to some extent in CESM, whereas they are slightly worse in ICON.

Validation against in-situ data suggests the near-surface cold biases to be exaggerated in the comparison with ERA5 and that the refined runs slightly outperform the unrefined runs (Table 2). The former is in line with recent studies that show that ERA5 is too warm in the Arctic (e.g., Batrak & Müller, 2019; Zampieri *et al.*, 2023). Based on an evaluation of the models, and ERA5, against measurements from 10 Arctic and 10 Antarctic stations, we find the refined runs to do better than the unrefined runs in most cases, although the improvements are predominantly small. Interestingly, both the refined and unrefined ICON simulations perform better than ERA5 in Antarctic summer for the performance metrics used here. Furthermore, the models are generally more accurate in the Arctic than the Antarctic and do better in summer than winter. Overall, ICON is the best model and MPAS the worst.

For the sea-ice covered regions, near-surface temperature biases can be affected by the representation of snow on sea ice (e.g., Batrak & Müller, 2019) and by the prescribed sea-ice thickness (e.g., Batrak & Müller, 2019; Zhang *et al.*, 2025). Over land, biases can be linked to the representation of the continental snow cover (Zhang *et al.*, 2025), and continental snow density and snow thermal conductivity (Simpson *et al.*, 2022; Zhang *et al.*, 2025).

Biases in hfls and hfss vary regionally with model-specific characteristics and notable refinement-induced improvements for certain regions (Figures 11–13). In many cases, the biases are associated with other near-surface variables, consistent with the bulk formulas, Equations (1)

and (2). Generally, both fluxes are overestimated over the NH and SH oceans during winter. An exception is CESM, which exhibits negative hfss biases over the North Atlantic, North Pacific, and SH oceans due to positive near-surface temperature biases. The positive hfss biases in CESM, on the other hand, are linked to near-surface winds over the North Atlantic, North Pacific, and SH ocean, and to overestimated rlds and cloudiness over the North Atlantic. In ICON and MPAS, positive hfss and hfss biases over the North Atlantic are associated with strong winds and cold, dry air, and to underestimated cloudiness in MPAS; over the North Pacific, the lack of consistency between biases in hfss and hfss and other fields suggests issues with the boundary-layer parametrizations. With refinement, the winter latent heat flux improves over the NH oceans in MPAS and over the SH ocean in MPAS and ICON.

Although we explored the cloud cover, radiative fluxes are also sensitive to cloud phase partitioning in both reanalysis (e.g., Bertossa *et al.*, 2025) and climate models (e.g., McIlhatten *et al.*, 2017, 2020). Underrepresented supercooled cloud liquid has been linked to too little Arctic rlds in CESM1 (McIlhatten *et al.*, 2017), whereas improvements in CESM2 (McIlhatten *et al.*, 2020) combined with the cloud microphysics bug fix (Zhu *et al.*, 2022) in our CESM experiments could be contributing to the smaller winter rlds biases over the Arctic (and Antarctic) compared with ICON and MPAS.

Refinement yields notable model-dependent effects over the Poles. Evaluating the refinement-induced normalized polar-cap differences for a wide range of fields reveals that MPAS exhibits the largest changes and ICON the smallest, underscoring that intermodel variability often surpasses differences between Arctic and Antarctic refinements in a single model. Our findings highlight the complexity of polar climate processes under refined resolution, with key open questions remaining regarding the mechanisms driving the model-specific cloud responses, and, in particular, the role of stratosphere–troposphere coupling in modulating large-scale circulation and surface climate impacts. More detailed analysis tailored for exploring these mechanisms is currently ongoing. Future work will also document and explore the influence of polar refinement under projected future climate change, using a complimentary coordinated multimodel experiment set based on the same three models as used here.

Given that we only have one realization for each experiment, the contribution of internal variability to the differences between the refined and unrefined experiments cannot be rigorously assessed. Using the spread across the 10 members from the CESM CMIP6 AMIP ensemble as a surrogate estimate of the internal variability reveals that the differences in general exceed the estimated internal

variability for temperature diagnostics, but not for zonal winds (not shown). However, the zonal wind changes in our experiments generally being in line with the changes in temperature supports attributing the changes to the refinement. Still, future studies should strive to include multiple realizations to further explore the role of internal variability on the influence of polar refinement.

The results are likely to be affected by differences in resolution between the models, both in the vertical and horizontal. In particular, though the very finest horizontal resolution is applied in similar regions in the three models, the latitudinal extent of the transition zone, where the horizontal resolution transitions between the finest and coarsest resolution, varies substantially (Figure 1). Future multimodel studies should strive to further harmonize the design of the refined and transitional regions.

ACKNOWLEDGEMENTS

We were supported by the European Commission H2020 project PolarRES (grant number 101003590). In addition, LSG was supported by the Research Council of Norway project BASIC (project number 325440). We acknowledge Sigma2, the National Infrastructure for High Performance Computing and Data Storage in Norway, for providing computing and storage resources for the MPAS runs (project NN8002K and NS8002K) and storage resources for joint data analysis and sharing (project NS8402K). Computing and data storage resources for the CESM simulations, including the Cheyenne supercomputer (<https://doi.org/10.5065/D6RX99HX>; Computational and Laboratory, 2019), were provided by the Computational and Information Systems Laboratory (CISL) at NSF-NCAR. The ICON runs were carried out on the Alfred Wegener Institute high-performance computing facility Albedo. This publication uses data from the WHOI OAFlux datasets supported by the NOAA's Global Ocean Monitoring and Observing Program and NASA's Making Earth System Data Records for Use in Research Environments (MEaSUREs) Program. The OAFlux and FLUXCOM data were retrieved from the International Land Model Benchmarking website (ILAMB, 2024).

CONFLICTS OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The model data are openly available through the Norwegian Infrastructure for Research Data (NIRD) Research Data Archive for the CESM experiments (<https://doi.org/10.11582/2025.HI2JQSG9>; Wijngaard, 2025) and the MPAS experiments (<https://doi.org/10.11582/2026.d0fxwfs0>; Mooney & Levine, 2026) and through the World

Data Center for Climate at the German Climate Computing Centre (Deutsches Klimarechenzentrum; DKRZ) for the ICON experiments (https://doi.org/10.26050/WDCC/PolarRES_ICON; Koehler & Handorf, 2025). Though the data used here were interpolated to coarse-resolution regular grids in the horizontal before analysis, data from the refined experiments was also interpolated to high-resolution regular grids and can be made available upon request.

ORCID

Lise Seland Graff  <https://orcid.org/0000-0003-3217-6329>

REFERENCES

- Bambach, N.E., Rhoades, A.M., Hatchett, B.J., Jones, A.D., Ullrich, P.A. & Zarzycki, C.M. (2022) Projecting climate change in South America using variable-resolution community Earth system model: An application to Chile. *International Journal of Climatology*, 42, 2514–2542. Available from: <https://doi.org/10.1002/joc.7379>
- Batrak, Y. & Müller, M. (2019) On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice. *Nature Communications*, 10, 4170. Available from: <https://doi.org/10.1038/s41467-019-11975-3>
- Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M.J. et al. (2008) Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, 134, 1337–1351. Available from: <https://doi.org/10.1002/qj.289>
- Bechtold, P., Semane, N., Lopez, P., Chaboureaud, J.-P., Beljaars, A. & Bormann, N. (2014) Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, 71, 734–753. Available from: <https://doi.org/10.1175/JAS-D-13-0163.1>
- Beljaars, A.C., Brown, A.R. & Wood, N. (2004) A new parametrization of turbulent orographic form drag. *Quarterly Journal of the Royal Meteorological Society*, 130, 1327–1347. Available from: <https://doi.org/10.1256/qj.03.73>
- Bertossa, C., L'Ecuyer, T., Henderson, D. & McIlhatten, E. (2025) Underrepresentation of ubiquitous opaque and transmissive arctic atmospheric states in modern reanalyses. *Journal of Climate*, 38, 4369–4392.
- Bintanja, R. & Andry, O. (2017) Towards a rain-dominated Arctic. *Nature Climate Change*, 7, 263–267. Available from: <https://doi.org/10.1038/nclimate3240>
- Bland, J., Gray, S., Methven, J. & Forbes, R. (2021) Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts. *Quarterly Journal of the Royal Meteorological Society*, 147, 3878–3898. Available from: <https://doi.org/10.1002/qj.4150>
- Bogenschutz, P.A., Gettelman, A., Morrison, H., Larson, V.E., Craig, C. & Schanen, D.P. (2013) Higher-order turbulence closure and its impact on climate simulations in the community atmosphere model. *Journal of Climate*, 26, 9655–9676. Available from: <https://doi.org/10.1175/JCLI-D-13-00075.1>
- Boza, B., Herrington, A., Ilicak, M., Danabasoglu, G. & Sen, O.L. (2025) Exploring climate of the euro-mediterranean using a variable-resolution configuration of the global community earth system model (VR-CESM). *Journal of Geophysical Research: Atmospheres*, 130, e2024JD041300. Available from: <https://doi.org/10.1029/2024JD041300>
- Bromwich, D.H., Ensign, A., Wang, S.-H. & Zou, X. (2024) Major artifacts in ERA5 2-m air temperature trends over antarctica prior to and during the modern satellite era. *Geophysical Research Letters*, 51, e2024GL111907. Available from: <https://doi.org/10.1029/2024GL111907>
- Butler, A.H., Thompson, D.W.J. & Heikes, R. (2010) The steady-state atmospheric circulation response to climate change-like thermal forcings in a simple general circulation model. *Journal of Climate*, 23, 3474–3496. Available from: <https://doi.org/10.1175/2010JCLI3228.1>
- Charlton-Perez, A.J., Baldwin, M.P., Birner, T., Black, R.X., Butler, A.H., Calvo, N. et al. (2013) On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, 118, 2494–2505. Available from: <https://doi.org/10.1002/jgrd.50125>
- Chen, F. & Dudhia, J. (2001) Coupling an advanced land surface–hydrology model with the penn state–ncar mm5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, 129, 569–585. Available from: [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2)
- Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V. et al. (1996) Modeling of land surface evaporation by four schemes and comparison with life observations. *Journal of Geophysical Research: Atmospheres*, 101, 7251–7268. Available from: <https://doi.org/10.1029/95JD02165>
- Cheng, Y., Hui, P., Liu, D., Fang, J., Wang, S., Wang, S. et al. (2023) MPAS-A variable-resolution simulations for summer monsoon over China: Comparison between global and regional configuration. *Journal of Geophysical Research: Atmospheres*, 128, e2022JD037541. Available from: <https://doi.org/10.1029/2022JD037541>
- Choi, H.-J. & Hong, S.-Y. (2015) An updated subgrid orographic parameterization for global atmospheric forecast models. *Journal of Geophysical Research: Atmospheres*, 120, 12445–12457. Available from: <https://doi.org/10.1002/2015JD024230>
- Computational and Laboratory, I. S. (2019) Cheyenne supercomputer. <https://doi.org/10.5065/D6RX99HX>
- Danabasoglu, G. (2019) NCAR CESM2 model output prepared for CMIP6 CMIP amip. <https://doi.org/10.22033/ESGF/CMIP6.7522>
- Danabasoglu, G., Lamarque, J.F., Bacmeister, J., Bailey, D.A., DuVivier, A.K., Edwards, J. et al. (2020) The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001916. Available from: <https://doi.org/10.1029/2019MS001916>
- Danielson, J.J. & Gesch, D.B. (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010). Technical report. <https://doi.org/10.3133/ofr20111073>
- Datta, R.T., Herrington, A., Lenaerts, J.T., Schneider, D.P., Trusel, L., Yin, Z. et al. (2023) Evaluating the impact of enhanced horizontal resolution over the Antarctic domain using a variable-resolution earth system model. *The Cryosphere*, 17, 3847–3866. Available from: <https://doi.org/10.5194/tc-17-3847-2023>
- Davy, R. & Outten, S. (2020) The arctic surface climate in CMIP6: status and developments since CMIP5. *Journal of Climate*, 33,

- 8047–8068. Available from: <https://doi.org/10.1175/JCLI-D-19-0990.1>
- Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M. et al. (2011) *A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization*. Offenbach, Germany: Deutscher Wetterdienst. Available from: https://www.cosmo-model.org/content/model/cosmo/coreDocumentation/cosmo_physics_4.20.pdf
- Eady, E.T. (1949) Long waves and cyclone waves. *Tellus*, 1, 33–52. Available from: <https://doi.org/10.3402/tellusa.v1i3.8507>
- Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V. et al. (2003) Implementation of noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research: Atmospheres*, 108, 2002JD003296. Available from: <https://doi.org/10.1029/2002JD003296>
- Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. et al. (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958. Available from: <https://doi.org/10.5194/gmd-9-1937-2016>
- Foken, T. (2008) *Micrometeorology*, Vol. 2. Berlin, Heidelberg: Springer. Available from: https://doi.org/10.1007/978-3-540-74666-9_6
- Fox-Rabinovitz, M., Cote, J., Dugas, B., Deque, M., McGregor, J.L. & Belochitski, A. (2008) Stretched-grid Model Intercomparison Project: decadal regional climate simulations with enhanced variable and uniform-resolution GCMs. *Meteorology and Atmospheric Physics*, 100, 159–178.
- Francelino, M.R., Schaefer, C., de Los Milagros Skansi, M., Colwell, S., Bromwich, D.H., Jones, P. et al. (2021) WMO Evaluation of Two Extreme High Temperatures Occurring in February 2020 for the Antarctic Peninsula Region. *Bulletin of the American Meteorological Society*, 102, E2053–E2061. Available from: <https://doi.org/10.1175/BAMS-D-21-0040.1>
- Gates, W.L., Boyle, J.S., Covey, C., Dease, C.G., Doutriaux, C.M., Drach, R.S. et al. (1999) An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, 80, 29–56. Available from: [https://doi.org/10.1175/1520-0477\(1999\)080<0029:AOOTRO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2)
- Gottelman, A., Callaghan, P., Larson, V., Zarzycki, C., Bacmeister, J., Lauritzen, P. et al. (2018) Regional climate simulations with the community earth system model. *Journal of Advances in Modeling Earth Systems*, 10, 1245–1265. Available from: <https://doi.org/10.1002/2017MS001227>
- Gottelman, A., Mills, M.J., Kinnison, D.E., Garcia, R.R., Smith, A.K., Marsh, D.R. et al. (2019a) the whole atmosphere community climate model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, 124, 12380–12403. Available from: <https://doi.org/10.1029/2019JD030943>
- Gottelman, A. & Morrison, H. (2015) Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, 28, 1268–1287. Available from: <https://doi.org/10.1175/JCLI-D-14-00102.1>
- Gottelman, A., Morrison, H., Thayer-Calder, K. & Zarzycki, C.M. (2019b) The impact of rimed ice hydrometeors on global and regional climate. *Journal of Advances in Modeling Earth Systems*, 11, 1543–1562. Available from: <https://doi.org/10.1029/2018MS001488>
- Goyal, R., Jucker, M., Sen Gupta, A., Hendon, H.H. & England, M.H. (2021) Zonal wave 3 pattern in the Southern Hemisphere generated by tropical convection. *Nature Geoscience*, 14, 732–738. Available from: <https://doi.org/10.1038/s41561-021-00811-3>
- Graff, L.S., Iversen, T., Bethke, I., Debernard, J.B., Seland, Ø., Bentsen, M. et al. (2019) Arctic amplification under global warming of 1.5 and 2 °C in NorESM1-Happi. *Earth System Dynamics*, 10, 569–598. Available from: <https://doi.org/10.5194/esd-10-569-2019>
- Hamilton, K., Wilson, R.J., Mahlman, J.D. & Umscheid, L.J. (1995) Climatology of the SKYHI troposphere–stratosphere–mesosphere general circulation model. *Journal of the Atmospheric Sciences*, 52, 5–43. Available from: [https://doi.org/10.1175/1520-0469\(1995\)052<0005:COTSTG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<0005:COTSTG>2.0.CO;2)
- Harris, L.M. & Lin, S.-J. (2013) A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, 141, 283–306. Available from: <https://doi.org/10.1175/MWR-D-11-00201.1>
- Harris, L.M. & Lin, S.-J. (2014) Global-to-regional nested grid climate simulations in the GFDL high resolution atmospheric model. *Journal of Climate*, 27, 4890–4910. Available from: <https://doi.org/10.1175/JCLI-D-13-00596.1>
- Harvey, B.J., Shaffrey, L.C. & Woollings, T.J. (2014) Equator-to-pole temperature differences and the extra-tropical storm track responses of the CMIP5 climate models. *Climate Dynamics*, 43, 1171–1182. Available from: <https://doi.org/10.1007/s00382-013-1883-9>
- Hastings, D.A., Dunbar, P.K., Elphinstone, G.M., Bootz, M., Murakami, H., Maruyama, H. et al. (1999) The global land one-kilometer base elevation (GLOBE) digital elevation model, version 1.0. *National Oceanic and Atmospheric Administration, National Geophysical Data Center*, 325, 80305–83328.
- Herrington, A.R., Lauritzen, P.H., Lofverstrom, M., Lipscomb, W.H., Gettelman, A. & Taylor, M.A. (2022) Impact of grids and dynamical cores in CESM2.2 on the surface mass balance of the Greenland ice sheet. *Journal of Advances in Modeling Earth Systems*, 14, e2022MS003192. Available from: <https://doi.org/10.1029/2022MS003192>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: <https://doi.org/10.1002/qj.3803>
- Hoesly, R.M., Smith, S.J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T. et al. (2018) Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, 11, 369–408. Available from: <https://doi.org/10.5194/gmd-11-369-2018>
- Hogan, R.J. & Bozzo, A. (2018) A flexible and efficient radiation scheme for the ECMWF model. *Journal of Advances in Modeling Earth Systems*, 10, 1990–2008. Available from: <https://doi.org/10.1029/2018MS001364>
- Holland, M.M. & Bitz, C.M. (2003) Polar amplification of climate change in coupled models. *Climate Dynamics*, 21, 221–232. Available from: <https://doi.org/10.1007/s00382-003-0332-6>
- Hollingsworth, A., Engelen, R.J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F. et al. (2008) Toward a monitoring and forecasting system for atmospheric composition: the gems project.

- Bulletin of the American Meteorological Society*, 89, 1147–1164. Available from: <https://doi.org/10.1175/2008BAMS2355.1>
- Hong, S.-Y. & Lim, J.-O.J. (2006) The WRF single-moment 6-class microphysics scheme (WSM6). *Asia-Pacific Journal of Atmospheric Sciences*, 42, 129–151.
- Hong, S.-Y. & Pan, H.-L. (1996) Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Monthly Weather Review*, 124, 2322–2339. Available from: [https://doi.org/10.1175/1520-0493\(1996\)124<2322:NBLVDI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2)
- Huang, X., Gettelman, A., Skamarock, W.C., Lauritzen, P.H., Curry, M., Herrington, A. et al. (2022) Advancing precipitation prediction using a new-generation storm-resolving model framework - SIMA-MPAS (V1.0): A case study over the western United States. *Geoscientific Model Development*, 15, 8135–8151. Available from: <https://doi.org/10.5194/gmd-15-8135-2022>
- Hunke, E.C., Lipscomb, W.H., Turner, A.K., Jeffery, N. & Elliott, S. (2015) *CICE: the Los Alamos sea ice model documentation and software user's manual version 5.1. Technical report LA-CC-06-012*. Los Alamos, New Mexico: Los Alamos National Laboratory. Available from: <https://github.com/CICE-Consortium/CICE-svn-trunk/releases/tag/cice-5.1.2>
- Huo, Y., Wang, H., Veneziani, M., Comeau, D., Osinski, R., Hillman, B.R. et al. (2025) E3sm-arctic: Regionally refined coupled model for advanced understanding of arctic systems interactions. *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004726. Available from: <https://doi.org/10.1029/2024MS004726>
- Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A. & Collins, W.D. (2008) Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, 113, 2008JD009944. Available from: <https://doi.org/10.1029/2008JD009944>
- ILAMB. (2024) International land model benchmarking: datasets. <https://www.ilamb.org/datasets.html> [Accessed 28th Nov 2025].
- Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M. et al. (2019) The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, 19, 3515–3556. Available from: <https://doi.org/10.5194/acp-19-3515-2019>
- Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H. et al. (2013) The MACC reanalysis: an 8 yr data set of atmospheric composition. *Atmospheric Chemistry and Physics*, 13, 4073–4109. Available from: <https://doi.org/10.5194/acp-13-4073-2013>
- Jiménez, P.A., Dudhia, J., González-Rouco, J.F., Navarro, J., Montávez, J.P. & García-Bustamante, E. (2012) A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, 140, 898–918. Available from: <https://doi.org/10.1175/MWR-D-11-00056.1>
- Juckes, M., Taylor, K.E., Durack, P.J., Lawrence, B., Mizielinski, M.S., Pament, A. et al. (2020) The CMIP6 data request (DREQ, version 01.00.31). *Geoscientific Model Development*, 13, 201–224. Available from: <https://doi.org/10.5194/gmd-13-201-2020>
- Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G. et al. (2019) The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, 6, 1–14. Available from: <https://doi.org/10.1038/s41597-019-0076-8>
- Klemp, J.B., Skamarock, W.C. & Dudhia, J. (2007) Conservative split-explicit time integration methods for the compressible non-hydrostatic equations. *Monthly Weather Review*, 135, 2897–2913. Available from: <https://doi.org/10.1175/MWR3440.1>
- Koehler, R. & Handorf, D. (2025) ERA5-based present-day ICON simulations with variable polar resolution. https://doi.org/10.26050/WDCC/PolarRES_ICON
- Köhler, R., Handorf, D., Jaiser, R., Dethloff, K., Zängl, G., Majewski, D. et al. (2021) Improved circulation in the northern hemisphere by adjusting gravity wave drag parameterizations in seasonal experiments with ICON-NWP. *Earth and Space Science*, 8, e2021EA001676. Available from: <https://doi.org/10.1029/2021EA001676>
- Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K. et al. (2024) The JRA-3Q reanalysis. *Journal of the Meteorological Society of Japan*, 102, 49–109. Available from: <https://doi.org/10.2151/jmsj.2024-004>
- Krinner, G., Genthon, C., Li, Z.-X. & Le Van, P. (1997) Studies of the antarctic climate with a stretched-grid general circulation model. *Journal of Geophysical Research: Atmospheres*, 102, 13731–13745. Available from: <https://doi.org/10.1029/96JD03356>
- Krinner, G., Llargeron, C., Ménégoz, M., Agosta, C. & Brutel-Vuilmet, C. (2014) Oceanic forcing of Antarctic climate change: A study using a stretched-grid atmospheric general circulation model. *Journal of Climate*, 27, 5786–5800. Available from: <https://doi.org/10.1175/JCLI-D-13-00367.1>
- Krinner, G., Magand, O., Simmonds, I., Genthon, C. & Dufresne, J.L. (2007) Simulated Antarctic precipitation and surface mass balance at the end of the twentieth and twenty-first centuries. *Climate Dynamics*, 28, 215–230. Available from: <https://doi.org/10.1007/s00382-006-0177-x>
- Lauritzen, P.H., Nair, R.D., Herrington, A.R., Callaghan, P., Goldhaber, S., Dennis, J.M. et al. (2018) NCAR release of CAM-SE in CESM2.0: a reformulation of the spectral element dynamical core in dry-mass vertical coordinates with comprehensive treatment of condensates and energy. *Journal of Advances in Modeling Earth Systems*, 10, 1537–1570. Available from: <https://doi.org/10.1029/2017MS001257>
- Lawrence, D.M., Fisher, R.A., Koven, C.D., Oleson, K.W., Swenson, S.C., Bonan, G. et al. (2019) The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11, 4245–4287. Available from: <https://doi.org/10.1029/2018MS001583>
- Lawrence, Z.D., Abalos, M., Ayarzagüena, B., Barriopedro, D., Butler, A.H., Calvo, N. et al. (2022) Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems. *Weather and Climate Dynamics*, 3, 977–1001. Available from: <https://doi.org/10.5194/wcd-3-977-2022>
- Li, C., Michel, C., Seland Graff, L., Bethke, I., Zappa, G., Bracegirdle, T.J. et al. (2018) Midlatitude atmospheric circulation responses under 1.5 and 2.0 °C warming and implications for regional impacts. *Earth System Dynamics*, 9, 359–382. Available from: <https://doi.org/10.5194/esd-9-359-2018>
- Liang, Y., Yang, B., Wang, M., Guo, Y., Sakaguchi, K. & Leung, L.R. (2023) Tropical cyclone strength, precipitation, and environment in variable resolution CAM-MPAS simulations over Western North Pacific. *Climate Dynamics*, 61, 2253–2267. Available from: <https://doi.org/10.1007/s00382-023-06677-y>

- Liang, Y., Yang, B., Wang, M., Tang, J., Sakaguchi, K., Leung, L.R. et al. (2021) Multiscale simulation of precipitation over East Asia by variable resolution CAM-MPAS. *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002656. Available from: <https://doi.org/10.1029/2021MS002656>
- Lin, M., Horowitz, L.W., Zhao, M., Harris, L., Ginoux, P., Dunne, J. et al. (2024) The GFDL variable-resolution global chemistry-climate model for research at the Nexus of US climate and air quality extremes. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS003984. Available from: <https://doi.org/10.1029/2023MS003984>
- Lindzen, R.S. & Farrell, B. (1980) A simple approximate result for the maximum growth rate of baroclinic instabilities. *Journal of the Atmospheric Sciences*, 37, 1648–1654. Available from: [https://doi.org/10.1175/1520-0469\(1980\)037<1648:ASARFT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1980)037<1648:ASARFT>2.0.CO;2)
- Liu, X., Ma, P.L., Wang, H., Tilmes, S., Singh, B., Easter, R.C. et al. (2016) Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of the community atmosphere model. *Geoscientific Model Development*, 9, 505–522. Available from: <https://doi.org/10.5194/gmd-9-505-2016>
- Lott, F. & Miller, M.J. (1997) A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, 123, 101–127. Available from: <https://doi.org/10.1002/qj.49712353704>
- Maoyi, M.L. & Abiodun, B.J. (2022) Investigating the response of the botswana high to El Niño Southern Oscillation using a variable resolution global climate model. *Theoretical and Applied Climatology*, 147, 1601–1615. Available from: <https://doi.org/10.1007/s00704-021-03892-7>
- McIlhatten, E.A., Kay, J.E. & L'Ecuyer, T.S. (2020) Arctic clouds and precipitation in the community earth system model version 2. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032521. Available from: <https://doi.org/10.1029/2020JD032521>
- McIlhatten, E.A., L'Ecuyer, T.S. & Miller, N.B. (2017) Observational evidence linking arctic supercooled liquid cloud biases in CESM to snowfall processes. *Journal of Climate*, 30, 4477–4495. Available from: <https://doi.org/10.1175/JCLI-D-16-0666.1>
- Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D.M. et al. (2017) Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, 10, 2057–2116. Available from: <https://doi.org/10.5194/gmd-10-2057-2017>
- Mi, X. & Liu, P. (2025) A comparative study of near-surface wind speed observations and reanalysis datasets in China over the past 38 years. *International Journal of Climatology*, 45, e8738. Available from: <https://doi.org/10.1002/joc.8738>
- Mironov, D., Ritter, B., Schulz, J.-P., Buchhold, M., Lange, M. & MacHulskaya, E. (2012) Parameterisation of sea and lake ice in numerical weather prediction models of the German Weather Service. *Tellus A: Dynamic Meteorology and Oceanography*, 64, 17330. Available from: <https://doi.org/10.3402/tellusa.v64i0.17330>
- Mooney, P. & Levine, X. (2026) PolarRES MPAS global simulations from 1985–2014 with an unrefined grid using ERA5 forcing data [Data set]. <https://doi.org/10.11582/2026.d0fxwfs0>
- Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G. et al. (2020) Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet. *Nature Geoscience*, 13, 132–137. Available from: <https://doi.org/10.1038/s41561-019-0510-8>
- Morlighem, M., Williams, C.N., Rignot, E., An, L., Arndt, J.E., Bamber, J.L. et al. (2017) BedMachine v3: complete bed topography and ocean bathymetry mapping of greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, 44, 11051–11061. Available from: <https://doi.org/10.1002/2017GL074954>
- Mouallem, J., Harris, L. & Benson, R. (2022) Multiple same-level and telescoping nesting in GFDL's dynamical core. *Geoscientific Model Development Discussions*, 2022, 1–28. Available from: <https://doi.org/10.5194/gmd-15-4355-2022>
- NASA/LARC/SD/ASDC. (2023) CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data in netCDF Edition 4.2. https://doi.org/10.5067/TERRA-AQUA-NOAA20/CERES/EBAF_L3B004.2
- Orr, A., Bechtold, P., Scinocca, J., Ern, M. & Janiskova, M. (2010) Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, 23, 5905–5926. Available from: <https://doi.org/10.1175/2010JCLI3490.1>
- Purich, A. & Doddridge, E.W. (2023) Record low Antarctic sea ice coverage indicates a new sea ice state. *Communications Earth & Environment*, 4, 314. Available from: <https://doi.org/10.1038/s43247-023-00961-9>
- Rahimi, S.R., Wu, C., Liu, X. & Brown, H. (2019) Exploring a variable-resolution approach for simulating regional climate over the Tibetan Plateau using VR-CESM. *Journal of Geophysical Research: Atmospheres*, 124, 4490–4513. Available from: <https://doi.org/10.1029/2018JD028925>
- Rantanen, M., Karpechko, A.Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K. et al. (2022) The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, 3, 1–10. Available from: <https://doi.org/10.1038/s43247-022-00498-3>
- Raschendorfer, M. (2001) The new turbulence parameterization of LM. COSMO Newsletter, No. 1, Consortium for Small-Scale Modeling, Offenbach, Germany, 89–97. <https://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter13/cnl13.pdf> [Accessed 28th Nov 2025].
- Rhoades, A.M., Ullrich, P.A., Zarzycki, C.M., Johansen, H., Margulis, S.A., Morrison, H. et al. (2018) Sensitivity of mountain hydroclimate simulations in variable-resolution CESM to microphysics and horizontal resolution. *Journal of Advances in Modeling Earth Systems*, 10, 1357–1380. Available from: <https://doi.org/10.1029/2018MS001326>
- Ringler, T.D., Thuburn, J., Klemp, J.B. & Skamarock, W.C. (2010) A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids. *Journal of Computational Physics*, 229, 3065–3090. Available from: <https://doi.org/10.1016/j.jcp.2009.12.007>
- Schmidt, F. (1977) Variable fine mesh in spectral global models. *Beiträge zur Physik der Atmosphäre*, 70, 211–217.
- Schrodin, R. & Heise, E. (2001) The multi-layer version of the DWD soil model TERRA_LM. <https://www.dwd.de/SharedDocs/downloads/DE/modelldokumentationen/nwv>

- [/veroeffentlichungen/Cosmo_tech_rep/costr2.pdf?__blob=publicationFilev=2](#) [Accessed 28th Nov 2025].
- Seifert, A. (2008) A revised cloud microphysical parameterization for COSMO-LME. *COSMO Newsletter*, 7, 25–28. Available from: https://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter07/cnl7_seifert.pdf [Accessed 28th Nov 2025].
- Shaw, J., McGraw, Z., Bruno, O., Storelvmo, T. & Hofer, S. (2022) Using satellite observations to evaluate model microphysical representation of arctic mixed-phase clouds. *Geophysical Research Letters*, 49, e2021GL096191. Available from: <https://doi.org/10.1029/2021GL096191>
- Simpson, I.R., Lawrence, D.M., Swenson, S.C., Hannay, C., McKinnon, K.A. & Truesdale, J.E. (2022) Improvements in winter-time surface temperature variability in the community earth system model version 2 (CESM2) related to the representation of snow density. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002880. Available from: <https://doi.org/10.1029/2021MS002880>
- Skamarock, W.C., Klemp, J.B., Duda, M.G., Fowler, L.D., Park, S.-H. & Ringler, T.D. (2012) A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering. *Monthly Weather Review*, 140, 3090–3105. Available from: <https://doi.org/10.1175/MWR-D-11-00215.1>
- Staniforth, A.N. & Mitchell, H.L. (1978) A variable-resolution finite-element technique for regional forecasting with the primitive equations. *Monthly Weather Review*, 106, 439–447. Available from: [https://doi.org/10.1175/1520-0493\(1978\)106<0439:AVRFET>2.0.CO;2](https://doi.org/10.1175/1520-0493(1978)106<0439:AVRFET>2.0.CO;2)
- Tang, Q., Klein, S.A., Xie, S., Lin, W., Golaz, J.-C., Roesler, E.L. et al. (2019) Regionally refined test bed in E3SM atmosphere model version 1 (EAMv1) and applications for high-resolution modeling. *Geoscientific Model Development*, 12, 2679–2706. Available from: <https://doi.org/10.5194/gmd-12-2679-2019>
- Taylor, P.C., Boeke, R.C., Boisvert, L.N., Feldl, N., Henry, M., Huang, Y. et al. (2022) Process Drivers, Inter-Model Spread, and the Path Forward: A Review of Amplified Arctic Warming. *Frontiers in Earth Science*, 9, 758361. Available from: <https://doi.org/10.3389/feart.2021.758361>
- Tegen, I., Hollrig, P., Chin, M., Fung, I., Jacob, D. & Penner, J. (1997) Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *Journal of Geophysical Research: Atmospheres*, 102, 23895–23915. Available from: <https://doi.org/10.1029/97JD01864>
- Thompson, D.W.J., Solomon, S., Kushner, P.J., England, M.H., Grise, K.M. & Karoly, D.J. (2011) Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nature Geoscience*, 5, 741–749. Available from: <https://doi.org/10.1038/ngeo1296>
- Thuburn, J., Ringler, T.D., Skamarock, W.C. & Klemp, J.B. (2009) Numerical representation of geostrophic modes on arbitrarily structured C-grids. *Journal of Computational Physics*, 228, 8321–8335. Available from: <https://doi.org/10.1016/j.jcp.2009.08.006>
- Tian, T., Yang, S., Lorentsen, J., Nielsen-Englyst, P. & Singha, S. (2024) Cooler Arctic surface temperatures simulated by climate models are closer to satellite-based data than the ERA5 reanalysis. *Communications Earth & Environment*, 5, 111. Available from: <https://doi.org/10.1038/s43247-024-01276-z>
- Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117, 1779–1800. Available from: [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSE>2.0.CO;2)
- Tramontana, G., Jung, M., Schwalm, C., Ichii, K., Camps-Valls, G., Raduly, B. et al. (2016) Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, 13, 4291–4313. Available from: <https://doi.org/10.5194/bg-13-4291-2016>
- Ullrich, P. (2014) SquadGen: spherical quadrilateral grid generator. <https://climtae.ucdavis.edu/squadgen.php> [Accessed 28th Nov 2025].
- van Kampenhout, L., Rhoades, A.M., Herrington, A.R., Zarzycki, C.M., Lenaerts, J.T., Sacks, W.J. et al. (2019) Regional grid refinement in an Earth system model: impacts on the simulated Greenland surface mass balance. *The Cryosphere*, 13, 1547–1564. Available from: <https://doi.org/10.5194/tc-13-1547-2019>
- van Marle, M.J.E., Kloster, S., Magi, B.I., Marlon, J.R., Daniau, A.-L., Field, R.D. et al. (2017) Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, 10, 3329–3357. Available from: <https://doi.org/10.5194/gmd-10-3329-2017>
- Waling, A., Herrington, A., Duderstadt, K., Dibb, J. & Burakowski, E. (2024) Using variable-resolution grids to model precipitation from atmospheric rivers around the Greenland ice sheet. *Weather and Climate Dynamics*, 5, 1117–1135. Available from: <https://doi.org/10.5194/wcd-5-1117-2024>
- Wang, W., Liu, X., Lin, G. & Wu, C. (2024) The impact of model horizontal resolution on simulating regional climate over East Asia using variable-resolution CESM2. *Journal of Geophysical Research: Atmospheres*, 129, e2023JD040308. Available from: <https://doi.org/10.1029/2023JD040308>
- Wicker, L.J. & Skamarock, W.C. (2002) Time-splitting methods for elastic models using forward time scheme. *Monthly Weather Review*, 130, 2088–2097. Available from: [https://doi.org/10.1175/1520-0493\(2002\)130<2088:TSMFEM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2)
- Wijngaard, R.R. (2025) VR-CESM2 model data for PolarRES WP2 runs: PD_ERA_UNREF and PD_ERA_REF_BIPOLAR. <https://doi.org/10.11582/2025.HI2JQSG9>
- Wijngaard, R.R., Herrington, A.R., Lipscomb, W.H., Leguy, G.R. & An, S.I. (2023) Exploring the ability of the variable-resolution Community Earth System Model to simulate cryospheric-hydrological variables in High Mountain Asia. *The Cryosphere*, 17, 3803–3828. Available from: <https://doi.org/10.5194/tc-17-3803-2023>
- Wijngaard, R.R., van de Berg, W.J., van Dalum, C.T., Herrington, A.R. & Levine, X.J. (2025) Arctic temperature and precipitation extremes in present-day and future storyline-based variable resolution community earth system model simulations. *Weather and Climate Dynamics*, 6, 1241–1266. Available from: <https://doi.org/10.5194/wcd-6-1241-2025>
- Wills, R.C.J., Herrington, A.R., Simpson, I.R. & Battisti, D.S. (2024) Resolving weather fronts increases the large-scale circulation response to gulf stream SST anomalies in variable-resolution CESM2 simulations. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004123. Available from: <https://doi.org/10.1029/2023MS004123>

- Wu, C., Liu, X., Lin, Z., Rhoades, A.M., Ullrich, P.A., Zarzycki, C.M. et al. (2017) Exploring a variable-resolution approach for simulating regional climate in the rocky mountain region using the VR-CESM. *Journal of Geophysical Research: Atmospheres*, 122, 10–939. Available from: <https://doi.org/10.1002/2017JD027008>
- Xie, A., Zhu, J., Qin, X. & Wang, S. (2023) The antarctic amplification based on modis land surface temperature and era5. *Remote Sensing*, 15, 3540. Available from: <https://doi.org/10.3390/rs15143540>
- Yu, L., Jin, X. & Weller, R.A. (2008) Multidecade global flux datasets from the objectively analyzed air-sea fluxes (oafux) project: latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables. In: *OAFux project technical report (OA-2008-01)*. Woods Hole, MA: Woods Hole Oceanographic Institution. Available from: https://oafux.whoi.edu/wp-content/uploads/sites/31/2020/06/OAFux_TechReport_3rd_release.pfd [Accessed 28th Nov 2025].
- Zampieri, L., Arduini, G., Holland, M., Keeley, S.P.E., Mogensen, K., Shupe, M.D. et al. (2023) A machine learning correction model of the winter clear-sky temperature bias over the arctic sea ice in atmospheric reanalyses. *Monthly Weather Review*, 151, 1443–1458. Available from: <https://doi.org/10.1175/MWR-D-22-0130.1>
- Zängl, G., Reinert, D. & Prill, F. (2022) Grid refinement in ICON v2. 6.4. *Geoscientific Model Development Discussions*, 2022, 1–32. Available from: <https://doi.org/10.5194/gmd-15-7153-2022>
- Zängl, G., Reinert, D., Rípodas, P. & Baldauf, M. (2015) The ICON (ICOSahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141, 563–579. Available from: <https://doi.org/10.1002/qj.2378>
- Zarzycki, C.M., Levy, M.N., Jablonowski, C., Overfelt, J.R., Taylor, M.A. & Ullrich, P.A. (2014) Aquaplanet experiments using CAM's variable-resolution dynamical core. *Journal of Climate*, 27, 5481–5503. Available from: <https://doi.org/10.1175/JCLI-D-14-00004.1>
- Zhang, G.J. & McFarlane, N.A. (1995) Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate Centre general circulation model. *Atmosphere-Ocean*, 33, 407–446. Available from: <https://doi.org/10.1080/07055900.1995.9649539>
- Zhang, Z., He, C., Berner, J., Jaye, A., Barlage, M., Liu, C. et al. (2025) Extending MPAS-NoahMP model capability beyond weather timescale: assessing land-atmosphere interactions for future subseasonal-to-seasonal applications. *ESS Open Archive*, 1–30. Available from: <https://doi.org/10.22541/essoar.175767377.72255512/v1>
- Zhu, J., Otto-Bliesner, B.L., Brady, E.C., Gettelman, A., Bacmeister, J.T., Neale, R.B. et al. (2022) LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002776. Available from: <https://doi.org/10.1029/2021MS002776>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Graff, L.S., Handorf, D., Köhler, R., Levine, X., Wijngaard, R.R., Marshall, G.J. et al. (2026) A multimodel intercomparison study of variable-resolution global models with grid refinement over the Arctic and Antarctic. *Quarterly Journal of the Royal Meteorological Society*, 152:e70192. Available from: <https://doi.org/10.1002/qj.70192>