Impact of grids and dynamical cores in CESM2.2 on the surface mass balance of the Greenland Ice Sheet

Adam R. Herrington[1], Peter H. Lauritzen[1], Marcus Lofverstrom[2], William H. Lipscomb[1], Andrew Gettelman[1] and Mark A. Taylor[3]

[1]National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, Colorado, USA
[2]Department of Geosciences, University of Arizona, 1040 E. 4th Street, Tucson, Arizona USA
[3]Sandia National Laboratories, Albuquerque, New Mexico, USA

Key Points:

- The CESM2.2 release includes several enhancements to the spectral-element dynamical core, including two new Arctic refined mesh configurations.
- Uniform resolution grids degrade the surface mass balance of the Greenland Ice Sheet compared with equivalent low resolution latitude-longitude grids.
- The refined Arctic meshes substantially improve the surface mass balance over all low resolution grids.

Corresponding author: Adam R. Herrington, `aherring@ucar.edu`

Abstract

Six different configurations, a mixture of grids and atmospheric dynamical cores available in the Community Earth System Model, version 2.2 (CESM2.2), are evaluated for their skill in representing the climate of the Arctic and the surface mass balance of the Greenland Ice Sheet (GrIS). The conventional $1°-2°$ uniform resolution grids systematically overestimate both accumulation and ablation over the GrIS. Of these conventional grids, the latitude-longitude grids outperform the quasi-uniform unstructured grids owing to their higher degrees of freedom in representing the GrIS. Two Arctic-refined meshes, with $1/4°$ and $1/8°$ refinement over Greenland, are documented as newly supported configurations in CESM2.2. The Arctic meshes substantially improve the simulated clouds and precipitation rates in the Arctic. Over Greenland, these meshes skillfully represent accumulation and ablation processes, leading to a more realistic GrIS surface mass balance. As CESM is in the process of transitioning away from conventional latitude-longitude grids, these new Arctic refined meshes improve the representation of polar processes in CESM by recovering resolution lost in the transition to quasi-uniform grids.

1 Introduction

General Circulation Models (GCMs) are powerful tools for understanding the meteorology and climate of the high latitudes, which are among the most sensitive regions on Earth to global and environmental change. GCMs differ vastly in their numerical treatment of polar regions because of the so-called *pole-problem* (Williamson, 2007). The pole problem refers to numerical instability arising from the convergence of meridian lines into polar singularities on latitude-longitude grids (e.g., Figure 1a, hereafter referred to as *lat-lon* grids). Depending on the numerics, methods exist to suppress this instability, and lat-lon grids may be advantageous for polar processes by representing structures with finer resolution than elsewhere in the computational domain. With the recent trend towards globally uniform unstructured grids, any potential benefits of lat-lon grids in polar regions may be lost. In this study, we evaluate a number of grids and dynamical cores (hereafter referred to as *dycores*) available in the Community Earth System Model, version 2.2 (CESM2.2; Danabasoglu et al., 2020), including new variable-resolution grids, to understand their impacts on the simulated Arctic climate. We focus specifically on the climate and surface mass balance of the Greenland Ice Sheet.

In the 1970s, the pole problem was largely defeated through the adoption of efficient spectral transform methods in GCMs (see Williamson, 2007, and references therein). These methods transform grid point fields into a global, isotropic representation in wave space, where linear operators (e.g., horizontal derivatives) in the (truncated) equation set can be solved exactly. While spectral transform methods are still used today, local numerical methods have become desirable for their ability to run efficiently on massively parallel systems. The pole problem has thus re-emerged in contemporary climate models that use lat-lon grids, and some combination of reduced grids (modified lat-lon grids, with cells elongated in the longitudinal direction over the polar regions) and polar filters are necessary to ameliorate this numerical instability (Jablonowski & Williamson, 2011). Polar filters subdue the growth of unstable computational modes by applying additional damping to the numerical solution over polar regions. This damping reduces the effective resolution in polar regions such that the resolved scales are *approximately* the same everywhere on the grid. We emphasize *approximately*, since it is at least conceivable that marginal increases in effective resolution occur over polar region in lat-lon grids, despite polar filtering, since resolved waves can be represented with more grid points than at lower latitudes.

Dycores built on lat-lon grids have some advantages over unstructured grids. Lat-lon coordinate lines are orthogonal, and aligned with zonally symmetric circulations that characterize many large-scale features of Earth's atmosphere. Lauritzen et al. (2010) has experimented with rotating lat-lon models such that their coordinate lines no longer align with an idealized, zonally balanced circulation. For the finite-volume lat-lon dycore considered in this paper (hereafter *FV*), numerical errors were shown to be largest when the polar singularity is rotated into the baroclinic zone (45°N latitude), generating spurious wave growth much earlier in the simulation than for other rotation angles. This illustrates the advantages of coordinate surfaces aligned with latitude bands, albeit an extreme example where the polar singularity and the polar filter are also contributing to the spurious wave growth. The unstructured grids all generate spurious baroclinic waves earlier on in the simulations than the (unrotated) lat-lon models, although the unstructured model considered in this paper, the spectral-element dycore (hereafter *SE*), holds a balanced zonal flow without spurious wave growth appreciably longer than the rotated FV experiments (Lauritzen et al., 2010). And unlike FV, the SE dycore has the same error characteristics regardless of how the grid is rotated.

The polar filter in the FV model impedes efficiency at large processor (CPU) counts because it requires a spectral transform, which have large communication overhead (Suarez & Takacs, 1995; Dennis et al., 2012). Unstructured grids support quasi-uniform grid spacing globally, and there is no pole-problem (e.g., Figure 1c). Conversely, unstructured grids are becoming increasingly common due to their improved performance on massively parallel systems and lack of constraints on grid structure (Taylor et al., 1997; Putman & Lin, 2007; Wan et al., 2013). This grid flexibility allows for the adoption of variable-resolution grids (e.g., Figure 2; hereafter abbreviated as *VR*), sometimes referred to as regional grid refinement. In principle, grid refinement over polar regions can make up for any loss of resolution in transitioning away from lat-lon grids (e.g., Figure 2). However, local grid refinement comes at the cost of a smaller CFL-limited time step in the refined region; the CFL-condition — short for Courant–Friedrichs–Lewy condition — is a necessary condition for numerical stability when using discrete data in time and space.

It is important to emphasize that the pole-problem is a distinctive feature of the dycore in atmospheric models. Polar filters do not directly interfere with the physical parameterizations, nor do they have any bearing on the surface models; e.g., the land model can take full advantage of the greater number of grid cells in polar regions on lat-lon grids. This is particularly relevant for the surface mass balance of the Greenland Ice Sheet (*SMB*; the integrated sum of precipitation and runoff), which relies on hydrological processes represented in the land model.

The SMB of the Greenland Ice Sheet (hereafter *GrIS*) is determined by processes occurring over a range of scales that are difficult to represent in GCMs (Pollard, 2010). GrIS precipitation is concentrated at the ice-sheet margins, where steep topographic slopes drive orographic precipitation. The truncated topography used by low resolution GCMs enables moisture to penetrate well into the GrIS interior, manifesting as a positive precipitation bias (Pollard & Groups, 2000; van Kampenhout et al., 2018). GrIS ablation areas (marginal regions where seasonal melting exceeds the annual mass input from precipitation) are typically less than 100 km wide and are confined to low-lying areas or regions with low precipitation. These narrow ablation zones are not fully resolved in low resolution GCMs, and may further degrade the simulated SMB. For example, CESM, version 2.0 (CESM2) underestimates ablation in the northern GrIS, leading to unrealistic ice advance when run with an interactive ice sheet component (Lofverstrom et al., 2020).

Regional climate models (RCMs) are commonly relied upon to provide more accurate SMB estimates. The limited area domain used by RCMs permits the use of high resolution grids, capable of resolving SMB processes, that can skillfully simulate the GrIS SMB (Box et al., 2004; Rae et al., 2012; Van Angelen et al., 2012; Fettweis et al., 2013; Mottram et al., 2017; Noël et al., 2018). However, unlike GCMs, RCMs are not a freely evolving system and the atmospheric state must be prescribed at the lateral boundaries of the model domain. The inability of the RCM solution to influence larger-scale dynamics outside the RCM domain (due to the prescribed boundary conditions) severely limits this approach from properly representing the role of the GrIS in the climate system. In addition, the boundary conditions are derived from a separate host model, which introduces inconsistencies due to differences in model design between the host model and the RCM.

In order to retain the benefits of RCMs in a GCM, van Kampenhout et al. (2018) utilized the VR capabilities of the SE dycore in CESM, generating a grid where Greenland is represented with 1/4° resolution, and elsewhere with the more conventional 1° resolution. The simulated SMB compared favorably to the SMB from RCMs and observations. The VR approach is therefore emerging as a powerful tool for simulating and understanding the GrIS and its response to different forcing scenarios, in the freely evolving GCM framework.

The SE dycore has been included in the model since CESM, version 1, but has been under active development ever since. This includes the switch to a dry-mass vertical coordinate (Lauritzen et al., 2018) and incorporation of an accelerated multi-tracer transport scheme (Lauritzen et al., 2017), made available in CESM2. This paper documents several additional enhancements to the SE dycore as part of the release of CESM2.2. These include three new VR configurations (Figure 2), two Arctic meshes and a Contiguous United-States mesh (`CONUS`; featured in Pfister et al. (2020)). While there are dozens of published studies using VR in CESM (e.g., Zarzycki et al., 2014; Rhoades et al., 2016; Gettelman et al., 2017; Burakowski et al., 2019; Bambach et al., 2021), these studies either used development code or collaborated closely with model developers. CESM2.2 is the first code release that contains out of the box VR functionality in CESM.

This study compares the representation of Arctic regions using the SE and FV dycores in CESM2.2 (see description below), as these two dycores treat high latitudes (i.e., the pole problem) in different ways. Section 2 documents the grids, dycores, and physical parameterizations used in this study, and also describes the experiments, datasets, and evaluation methods. Section 3 analyzes the results of the experiments, and Section 4 provides a general discussion and conclusions.

2 Methods

2.1 Dynamical cores

The atmospheric component of CESM2.2 (Danabasoglu et al., 2020), the Community Atmosphere Model, version 6.3 (CAM6; Gettelman et al., 2019; Craig et al., 2021), supports several different atmospheric dynamical cores. These include dycores on lat-lon grids, such as finite-volume (FV; Lin, 2004) and Eulerian spectral transform (EUL; Collins et al., 2006) models, and dycores built on unstructured grids, including spectral-element (SE; Lauritzen et al., 2018) and finite-volume 3 (FV3; Putman & Lin, 2007) models. This study compares the performance of the SE and FV dycores, omitting the EUL and FV3 dycores. CESM2 runs submitted to the Coupled Model Intercomparison Project Phase 6 (Eyring et al., 2016) used the FV dycore, whereas the SE dycore is often used for global high-resolution simulations (e.g., Small et al., 2014; Bacmeister & Coauthors, 2018; Chang et al., 2020) due to its higher throughput on massively parallel systems (Dennis et al., 2012).

2.1.1 Finite-volume (FV) dynamical core

The FV dycore is a hydrostatic model that integrates the equations of motion using a finite-volume discretization on a spherical lat-lon grid (Lin & Rood, 1997). The 2D dynamics evolve in floating Lagrangian layers that are periodically mapped to an Eu-

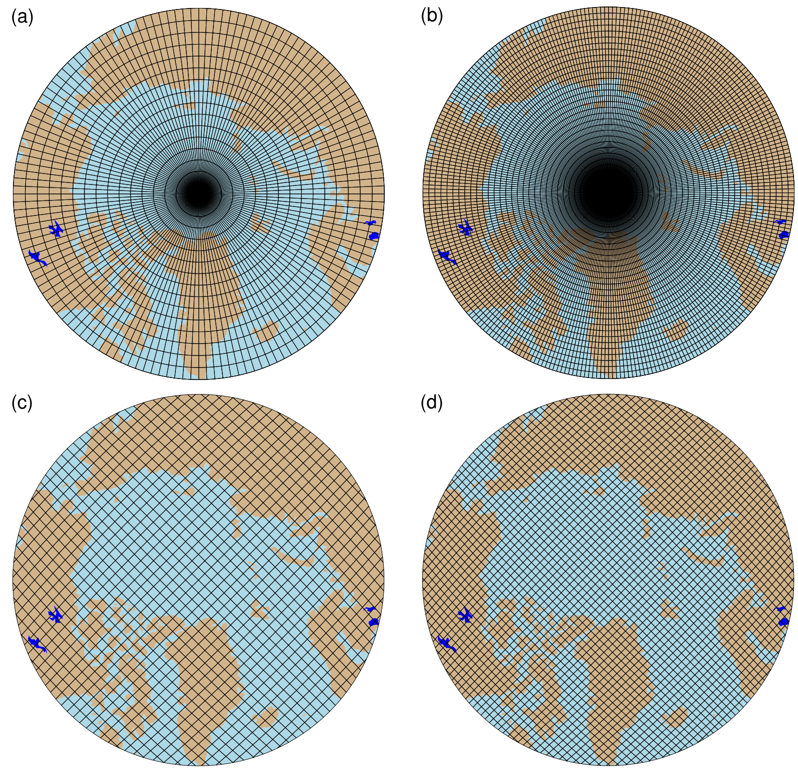

Figure 1. Computational grids for the uniform $1^\circ - 2^\circ$ grids in this study. Grids names after Table 1, (a) `f19`, (b) `f09`, (c) `ne30pg2` and (d) `ne30pg3`.

lerian reference grid in the vertical (Lin, 2004). Hyperviscous damping is applied to the divergent modes, and is increased in the top few layers (referred to as a *sponge layer*) to prevent undesirable interactions with the model top, such as wave reflection (Lauritzen et al., 2011). A polar filter damps computational instability due to the convergence of meridians, permitting a longer time step. It takes the form of a Fourier filter in the zonal direction, with the damping coefficients increasing monotonically in the meridional direction (Suarez & Takacs, 1995). The form of the filter is designed to slow down the propagation of large zonal wave-numbers to satisfy the CFL condition of the shortest resolved wave at some reference latitude.

2.1.2 Spectral-element (SE) dynamical core

The SE dycore is a hydrostatic model that integrates the equations of motion using a high-order continuous Galerkin method (Taylor et al., 1997; Taylor & Fournier, 2010). The computational domain is a cubed-sphere grid tiled with quadrilateral elements (see Figure 2). Each element contains a fourth-order basis set in each horizontal direction, with the solution defined at the roots of the basis functions, the Gauss-Lobatto-Legendre quadrature points. This results in 16 nodal points per element, with 12 of the points lying on the (shared) element boundary. Communication between elements uses the direct stiffness summation (Canuto et al., 2007), which applies a numerical flux to the element boundaries to reconcile overlapping nodal values and produce a continuous global basis set.

As with the FV dycore, the dynamics evolve in floating Lagrangian layers that are subsequently mapped to an Eulerian reference grid. A dry mass vertical coordinate was

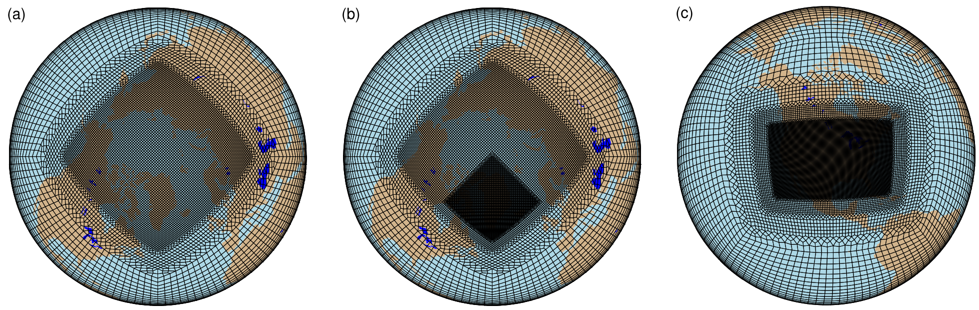

Figure 2. Spectral-element grid for the VR grids in this study, (a) `Arctic`, (b) `Arctic − GrIS` and (c) `CONUS`. Note that this is not the computational grid; each element has 3 × 3 independent grid points.

recently implemented for thermodynamic consistency with condensates (Lauritzen et al., 2018). The 2D dynamics have no implicit dissipation, and so hyperviscosity operators are applied to all prognostic variables to remove spurious numerical errors (Dennis et al., 2012). Laplacian damping is applied in the sponge layer.

The SE dycore supports regional grid refinement via its VR configuration, requiring two enhancements over uniform-resolution setups. Firstly, as the numerical viscosity increases with resolution, explicit hyperviscosity relaxes according to the local element size, reducing in strength by an order of magnitude per halving of the grid spacing. A tensor-hyperviscosity formulation is used (Guba et al., 2014), which adjusts the coefficients in two orthogonal directions to more accurately target highly distorted quadrilateral elements. Secondly, the topography boundary conditions are smoothed in a way that does not excite grid scale modes, and so the NCAR topography software (Lauritzen et al., 2015) has been modified to scale the smoothing radius by the local element size, resulting in rougher topography in the refinement zone.

For SE grids with quasi-uniform grid spacing, the SE tracer transport scheme is replaced with the Conservative Semi-Lagrangian Multi-tracer transport scheme (CSLAM) (Lauritzen et al., 2017). Atmospheric tracers have large, nearly discontinuous horizontal gradients that are difficult to represent with spectral methods, which are prone to oscillatory "Gibbs-ringing" errors (Rasch & Williamson, 1990). CSLAM has improved tracer property preservation and accelerated multi-tracer transport. It uses a separate grid from the spectral-element dynamics, dividing each element into 3×3 control volumes with quasi-equal area. The physical parameterizations are computed from the state on the CSLAM grid, which has clear advantages over the original SE dycore in which the physics are evaluated Gauss-Lobatto-Legendre points (Herrington et al., 2018).

2.2 Grids

We evaluate model simulations on six different grids in this study (Table 1). The FV dycore is run with nominal 1° and 2° grid spacing, referred to as `f09` and `f19`, respectively (Figure 1a,b). We also run the 1° equivalent of the SE-CSLAM grid, referred to as `ne30pg3` (Figure 1d), where ne refers to a grid with $ne\times ne$ elements per cubed-sphere face, and pg denotes that there are $pg\times pg$ control volumes per element for computing the physics. We run an additional 1° SE-CSLAM simulation with the physical parameterizations computed on a grid with 2×2 control volumes per element, `ne30pg2` (Figure 1c; Herrington et al., 2019, note CSLAM is still run on the 3 × 3 control volume grid).

grid name	dycore	Δx_{eq} (km)	$\Delta x_{\mathrm{refine}}$ (km)	Δt_{phys} (s)
`f19`	FV	278	-	1800
`f09`	FV	139	-	1800
`ne30pg2`	SE-CSLAM	167	-	1800
`ne30pg3`	SE-CSLAM	111	-	1800
`ne30pg3`*	SE-CSLAM	111	-	450
`Arctic`	SE	111	28	450
`Arctic − GrIS`	SE	111	14	225

Table 1. Grids and dycores used in this study. Δx_{eq} is the average equatorial grid spacing, $\Delta x_{\mathrm{refine}}$ is the grid spacing in the refined region (if applicable), and Δt_{phys} is the physics time step. FV refers to the finite-volume dycore, SE the spectral-element dycore, and SE-CSLAM the spectral-element dycore with CSLAM tracer advection. We use the `ne30pg3` grid for two runs with different values of Δt_{phys}.

Three VR meshes were developed for the CESM2.2 release to support grid refinement over the Arctic and the United States (Figure 2). This paper serves as the official documentation of these grids. The VR meshes were developed using the software package SQuadgen (`https://github.com/ClimateGlobalChange/squadgen`). The `Arctic` grid is a 1° grid with 1/4° regional refinement over the broader Arctic region. The `Arctic−GrIS` grid is identical to the `Arctic` grid, but with an additional patch covering the island of Greenland with 1/8° resolution. The `CONUS` grid contains 1/8° refinement over the United States, and 1° everywhere else. The `CONUS` grid is not discussed any further in this paper; see Pfister et al. (2020) for simulations with the `CONUS` grid.

The accuracy of the simulated surface mass balance is expected to be sensitive to grid resolution. Figure 3b shows the average grid spacing over the Greenland Ice Sheet (*GrIS* hereafter) in all six grids in this study. The `ne30pg2` grid has the coarsest representation with an average grid spacing (Δx) of $\Delta x = 160$ km, and the `Arctic−GrIS` grid has the highest resolution with an average grid spacing of $\Delta x = 14.6$ km, similar to the 11 km grid spacing of the RACMO2.3 grid. The `ne30pg3` grid has an average $\Delta x =$ 111.2 km, substantially coarser than the `f09` grid, with an average $\Delta x = 60$ km. Although `ne30pg3` and `f09` have similar average grid spacing over the entire globe, and comparable computational costs, the convergence of meridians on the FV grid enhances the resolution over the GrIS. The `Arctic` grid has an average grid spacing of $\Delta x = 27.8$ km, and is about 10 times more expensive than the 1° models. The `Arctic−GrIS` grid is about twice as expensive as the `Arctic` grid.

The physics time step depends on the grid resolution. Increased horizontal resolution permits faster vertical velocities that reduce characteristic time scales, so the physics time step is reduced to avoid large time truncation errors (Herrington & Reed, 2018). The `Arctic` and `Arctic − GrIS` grids are run with a 4× and 8× reduction in physics time step relative to the default 1800 s time step used in the 1° and 2° grids (Table 1).

All grids and dycores in this study use 32 hybrid pressure-sigma levels in the vertical, with a model top of 2 hPa or about 40 km. However, note that any grid or dycore can in principle be run with a higher model top or finer vertical resolution.

2.3 Physical parameterizations

All simulations in this study use the CAM6 physical parameterization package (hereafter referred to as the *physics*; Gettelman et al., 2019). The physics in CAM6 differs from its predecessors through the incorporation of high-order turbulence closure, Cloud Layers Unified by Binormals (CLUBB; Golaz et al., 2002; Bogenschutz et al., 2013), which jointly acts as a planetary boundary layer, shallow convection, and cloud macrophysics scheme. CLUBB is coupled with the MG2 microphysics scheme (Gettelman & Morrison, 2015; Gettelman et al., 2015), which computes prognostic precipitation and uses classical nucleation theory to represent cloud ice for improved cloud-aerosol interactions. Deep convection is parameterized using a convective quasi-equilibrium, mass flux scheme (Zhang & McFarlane, 1995; Neale et al., 2008) and includes convective momentum transport (Richter et al., 2010). Boundary layer form drag is modeled after Beljaars et al. (2004), and orographic gravity wave drag is represented with an anisotropic method informed by the orientation of topographic ridges at the sub-grid scale (the ridge orientation is derived from a high-resolution, global topography dataset (J. J. Danielson & Gesch, 2011)).

Initial simulations with the `ne30pg3` SE grid produced weaker shortwave cloud forcing relative to the tuned finite-volume dycore in the standard CESM2 configuration. The SE dycore in CESM2.2 therefore has two CLUBB parameter changes to provide more realistic cloud forcing and top-of-atmosphere radiation balance. We reduced the width of the sub-grid distribution of vertical velocity (`clubb_gamma` $=$ `0.308` $\rightarrow$ `0.270`) and also reduced the strength of the damping for horizontal component of turbulent energy (`clubb_c14` $=$ `2.2` $\rightarrow$ `1.6`) to increase cloudiness. For a description of how CLUBB parameters impact the simulated climate, see Guo et al. (2015).

2.4 Simulated surface mass balance (SMB)

All grids and dycores simulate the GrIS SMB, which is the sum of mass accumulation of precipitation and mass loss from ablation. The latter is the sum of evaporation, sublimation and liquid runoff, with runoff being a combination of liquid precipitation and snow and ice melt. Not all liquid precipitation or snow/ice melt runs off the ice sheet; this water can penetrate pore spaces in the snowpack/firn layer and freeze, increasing the ice mass. These relevant SMB processes are represented by different CESM components, but it is the Community Land Model, version 5 (CLM; Lawrence et al., 2019), that aggregates these processes and computes the SMB.

CLM runs on the same grid as the atmosphere, and uses a downscaling technique to account for sub-grid variability in SMB. In short, the ice sheet patch in a CLM grid cell is subdivided into 10 elevation classes (ECs), each with a distinct surface energy balance and SMB. The area fraction of each EC is derived from a high-resolution GrIS elevation dataset. The near-surface air temperature, humidity, and air density are calculated for each EC using an assumed lapse rate and the elevation difference from the grid-cell mean. Precipitation from CAM is repartitioned into solid or liquid based on the surface temperature of the EC; precipitation falls as snow for temperatures between $T < -2^\circ$ C, as rain for $T > 0^\circ$ C, and as a linear combination of rain and snow for temperatures between -2° C and 0° C. Snow accumulation in each EC is limited to a depth of 10 m liquid water equivalent. Any snow above the 10 m cap contributes towards ice accumulation in the SMB. Refreezing of liquid water within the snowpack is an additional source of ice. Integrating over all ECs, weighting by the area fractions, provides a more accurate SMB than would be found using the grid-cell mean elevation. For a more detailed description of how the SMB is computed in CESM, we refer the reader to Lipscomb et al. (2013); Sellevold et al. (2019); van Kampenhout et al. (2020); Muntjewerf et al. (2021).

Changes in ice depth, but not snow depth, count toward the SMB. That is, snow accumulation above the 10 m cap contributes a positive SMB, and surface ice melting (after melting of the overlying snow) yields a negative SMB. Since snow in the accumulation zone must reach the cap to simulate a positive SMB, the snow depths on the VR grids were spun up by forcing CLM in standalone mode, cycling over data from a 20-year `Arctic` `FHIST` simulation (a model simulation with prescribed, observed sea-surface conditions) for about 500 years. The uniform-resolution grids are initialized with the SMB

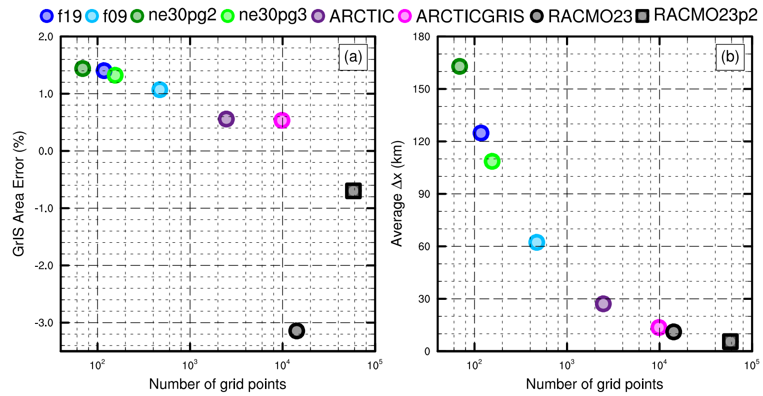

Figure 3. The spatial properties of the GrIS as represented by different grids in this study. (Left) GrIS area error, computed as the relative differences from a 4-km dataset used to create the CESM ice masks, (right) approximate average grid spacing over GrIS.

from an existing `f09` spun-up initial condition. In the simulations described here, the GrIS is prescribed at its observed, modern extent and thickness.

2.5 SMB Analysis

We seek to integrate SMB components over a GrIS ice mask and to diagnose their contributions to the GrIS mass budget. However, the ice masks vary across the grids, especially in comparison to the RACMO3.2 ice mask, whose total area is about 3% less than that of the reference dataset (Figure 3). CLM's dataset creation tool generates the model ice mask by mapping a high-resolution dataset to the target grid using the Earth System Modeling Framework (ESMF) first-order conservative remapping algorithm (Team et al., 2021). The figure suggests that the mapping errors are less than 1.5% across the CESM2 grids. The area errors in Figure 3 may seem small, but even $1-2\%$ area differences can lead to large differences in integrated SMB (Hansen et al., 2022).

We have taken a common-ice-mask approach by mapping all model fields to the lowest-resolution grids, i.e., the `f19` and `ne30pg2` grids, and integrating over these low-resolution ice masks. The use of low-resolution common ice masks is a conservative decision, and is justified because we seek to use first-order remapping algorithms to map fields to the common ice mask, which is not generally reliable when mapping to a higher-resolution grid than the source grid. We use two remapping algorithms: ESMF first-order conservative and the TempestRemap (Ullrich & Taylor, 2015) high-order monotone algorithm. Since mapping errors are sensitive to grid type, we evaluate all quantities on both common ice masks, the `f19` and `ne30pg2` masks. Thus, we evaluate an integrated quantity on a given grid up to four times to estimate the uncertainty due to differences in grid type and remapping algorithms.

The SMB is expressed in a form that is agnostic of water phase, a total water mass balance, to facilitate comparisons across different grids with different ice masks and to increase consistency with the variables available in the RACMO datasets. The SMB for total water can be expressed as:

$$SMB = accumulation + runoff + evaporation/sublimation, \tag{1}$$

where all terms have consistent sign conventions (positive values contribute mass, and negative values reduce mass). The accumulation source term refers to the combined solid and liquid precipitation, runoff refers to the liquid water sink, and evaporation/sublimation is the vapor sink. Since the runoff term aggregates many processes, we isolate the melting contribution by also tracking the combined melt of snow and ice. Note that this SMB expression is different from the internally computed SMB described in the previous section.

We consider two approaches for mapping and integrating the SMB components over the common ice masks:

1. Map the grid-cell mean quantities to the common grid, and integrate the mapped fields over the common ice masks.
2. Map the patch-level quantities (i.e., the state over the ice fractional component of the grid cell) to the common grid, and integrate the mapped fields over the common ice masks.

Note that we are mapping to low-resolution grids that have larger GrIS areas than the source grids (Figure 3). Since the components of equation 1 are not confined to the ice mask, method 1 reconstructs the SMB over the portion of the common ice mask that is outside the ice mask on the source grid. While this may be a an acceptable way to reconstruct the mass source terms over different ice masks, ice melt is zero outside the source ice mask, and so method 1 will underestimate the mass sink term. This underestimation is systematic in method 2, where all variables are exclusive to the ice mask; mapping to a lower-resolution grid will dilute a field of non-zero values over the ice mask with a field of zeros outside the ice mask. However, patch-level values for processes exclusive to the ice mask (e.g., ice melt) will on average have larger magnitudes than the the grid-mean quantities used in method 1.

The different error characteristics of the two methods are used to diversify the ensemble. Each of the four regridding combinations (with conservative and high-order remapping to the `f09` and `ne30pg2` grids) are repeated with each method, resulting in (up to) eight values for each integrated quantity. Unfortunately, the patch-level values of evaporation/sublimation are not available from the model output, and we estimate their contribution by zeroing out the field for grid cells that have no ice, prior to mapping to the common ice mask. This will degrade the SMB estimates using method 2, however we are more interested in characterizing the behavior of individual processes across grids and dycores, expressed as the components of the SMB, rather than the SMB itself.

2.6 Experimental design

All simulations described here use an identical transient 1979-1998 Atmospheric Model Inter-comparison Project (AMIP) configuration, with prescribed monthly sea-surface temperature and sea ice following Hurrell et al. (2008). In CESM terminology, AMIP simulations use the `FHIST` computational set and run out of the box in CESM2.2.

2.7 Observational Datasets

We use several observational datasets (Table 2) to assess the performance of the simulations. SMB datasets are gathered from multiple sources. Regional Atmospheric Climate Model, version 2.3 11km (RACMO23; Noël et al., 2015) and version 2.3p2 5.5km (RACMO2.3p2; Noël et al., 2018, 2019) are RCM simulations targeting Greenland, forced by ERA renalyses products at the domain's lateral boundaries. The RACMO simulations have been shown to perform skillfully against observations and are often used as modeling targets (e.g., Evans et al., 2019; van Kampenhout et al., 2020).

In-situ SMB (snow pit and ice cores) and radar accumulation datasets (e.g., IceBridge) are maintained in The Land Ice Verification and Validation toolkit (LIVVkit), version 2.1 (Evans et al., 2019). However, these point-wise measurements are difficult

data product	years used in this study	resolution	citation
ERA5	1979-1998	1/4°	Copernicus (2019)
CERES-EBAF ED4.1	2003-2020	1°	Loeb et al. (2018)
CALIPSO-GOCCP	2006-2017	1°	Chepfer et al. (2010)
RACMO2.3	1979-1998	11 km	Noël et al. (2015)
RACMO2.3p2	1979-1998	5.5 km	Noël et al. (2019)

Table 2. Description of observational datasets used in this study.

to compare to model output spanning several different grids, especially since the SMB from each elevation class is not available from the model output. We used a nearest-neighbor technique for an initial analysis, which showed that the model biases are similar to those computed using the RACMO datasets. Because of the uncertainty of comparing gridded fields to point-wise measurements, and the lack of information added with regard to model biases, we omitted these datasets from our analysis.

3 Results

3.1 Tropospheric temperatures

Before delving into the simulated Arctic climate conditions, we assess the global mean differences between the various grids and dycores. Figure 4 shows 1979-1998 annual mean, zonal mean height plots expressed as differences between uniform-resolution grids and dycores. The `f09` grid is warmer than the `f19` grid, primarily in the mid-to-high latitudes throughout the depth of the troposphere. This is a common response to increasing horizontal resolution in GCMs (Pope & Stratton, 2002; Roeckner et al., 2006). Herrington and Reed (2020) have shown that this occurs in CAM due to higher resolved vertical velocities which, in turn, generate more condensational heating in the CLUBB macrophyiscs. The right column in Figure 4a supports this interpretation, showing an increase in the climatological CLUBB heating at all latitudes in the `f09` grid, but with the largest increase in the mid-latitudes.

As the SE dycore is less diffusive than the FV dycore, the resolved vertical velocities are larger in the SE dycore, and so the `ne30pg3` troposphere is modestly warmer than `f09` (Figure 4b). The stratosphere responds differently, with `ne30pg3` much cooler than `f09` in the mid-to-high latitudes. Figure 4c also shows small temperature differences between `ne30pg3` and `ne30pg2`, with `ne30pg3` slightly warmer near the tropopause at high latitudes. This is consistent with the similar climates found for these two grids by Herrington et al. (2019).

Comparing the VR grids to the uniform-resolution grids is complicated because we simultaneously increase the resolution and reduce the physics time-step, both of which influence the solution (Williamson, 2008). We therefore run an additional `ne30pg3` simulation with the shorter physics time step used in the `Arctic` grid (450 s), referred to as `ne30pg3`* (Table 1). Figure 5a shows the difference between `ne30pg3`* and `ne30pg3` for climatological summer temperatures in zonal-mean height space. The troposphere is warmer with the reduced time step, and the mechanism is similar in that the shorter time step increases resolved vertical velocities (not shown) and CLUBB heating (right panel in Figure 5a). Figure 5b shows the difference in climatological summer temperature between the `Arctic` grid and the `ne30pg3`* grid. With the same physics time step, the greater condensational heating and warmer temperatures are confined to the refined Arctic region.

Figure 5c shows that the `Arctic−GrIS` grid is much warmer than the `Arctic` grid in the Arctic summer. This may be due, in part, to the shorter physics time step, but

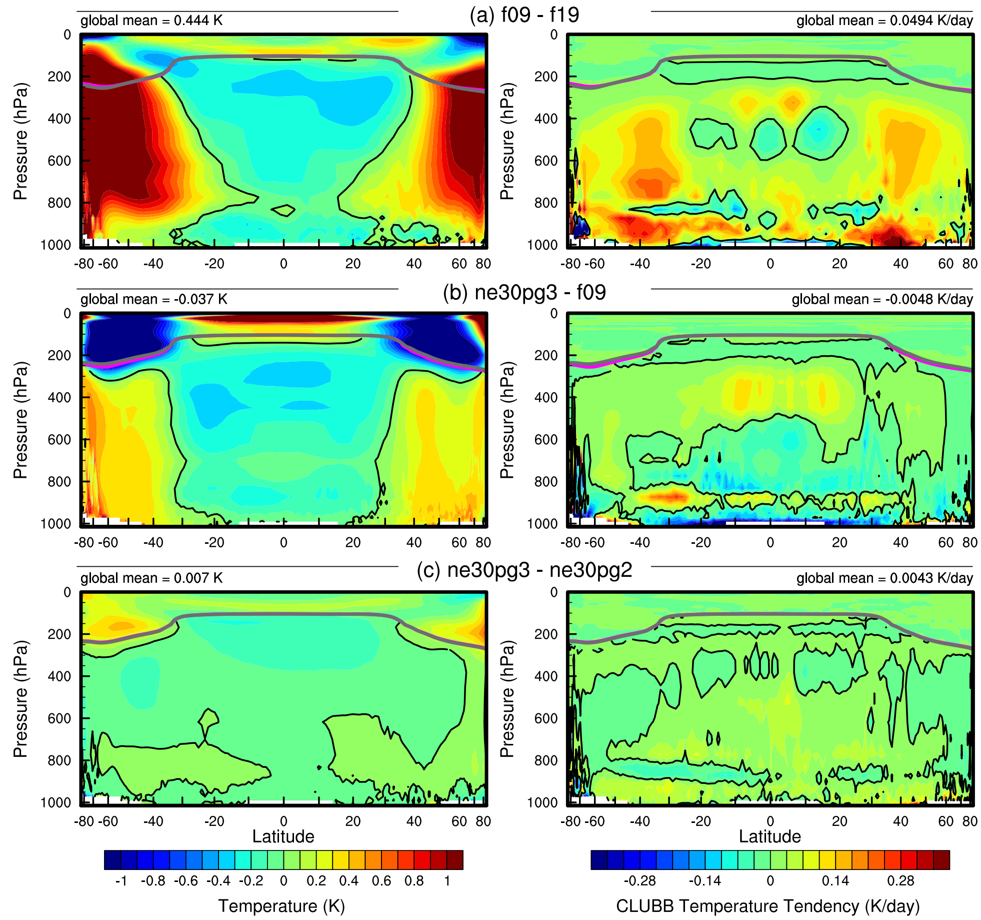

Figure 4. 1979-1998 annual mean temperature (left column) and CLUBB temperature tendencies (right column) in zonal mean height space, expressed as differences between the various $1^\circ - 2^\circ$ grids. The thick grey and magenta lines are the tropopause for the control run and the test run, respectively.

the temperature response is too large to be explained by the CLUBB changes alone. This summer warming appears to be a result of variations in the stationary wave pattern, with anomalous southerly winds to the west of Greenland (not shown). This dynamic response is interesting, because other than the physics time step, the only difference between the `Arctic−GrIS` and `Arctic` runs is the doubling of resolution over Greenland. This behavior will be explored further in a future study.

It is useful to understand summer temperature biases due to their control on ice and snow melt over the GrIS (Ohmura, 2001). Figure 6 shows the 1979–1998 lower troposphere summer temperature bias relative to ERA5, computed by equating a layer mean virtual temperature with the 500–1000 hPa geopotential thickness. The results are consistent with the zonal mean height plots; increasing resolution from `f19` to `f09` warms the climate, and the 1° SE grids are warmer than the FV grids. The FV summer temperatures are persistently colder than ERA5, whereas the 1° SE grids are not as cold, and are actually warmer than ERA5 at high-latitudes, north of 75°. All grids show a north-south gradient in bias over Greenland, with the summer temperature bias more positive for the northern part of the ice sheet. This pattern is also evident in the near surface temperature bias over Greenland (not shown).

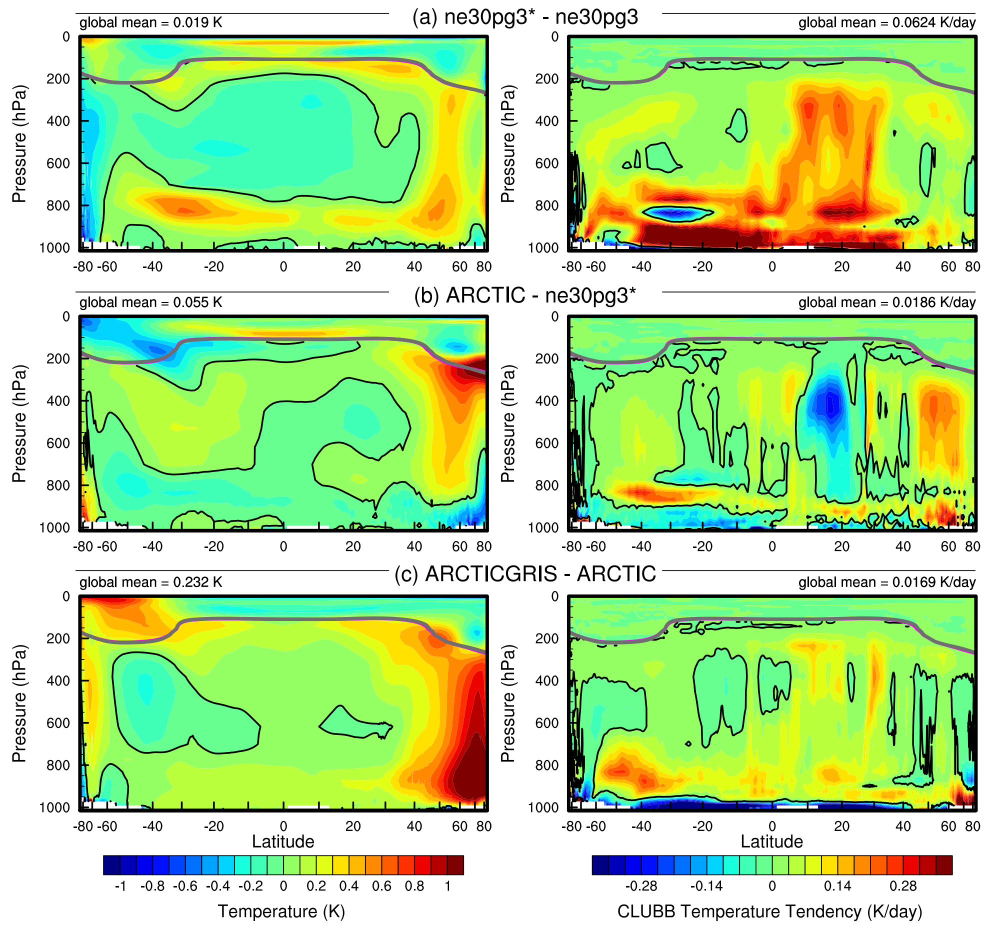

Figure 5. As in Figure 4 but for the short-time-step experiment and the VR grids. The fields plotted are for the climatological northern hemisphere summer. We focus on summer because that is when the resolution response is largest, and the refined regions are located in the northern hemisphere.

The `Arctic` grid has summer temperatures similar to the 1° SE grids, but is slightly warmer over northern Eurasia and the North Pole (Figure 6). An anomalous cooling patch forms to the west of Greenland, centered over Baffin Island. The `Arctic` − `GrIS` grid is warmer than the `Arctic` grid over most of the Arctic, but with a similar spatial pattern of summer temperature bias.

Some of these temperature differences may be related to different summer shortwave cloud forcing for the various grids and dycores. Figure 7 shows the summer shortwave cloud forcing bias in six runs, using the CERES product. A negative bias corresponds to excessive reflection and cooling. The uniform grids have similar biases, with the clouds reflecting 20–40 W/m^2 too much shortwave radiation over a wide swath of the Arctic, primarily the land masses. There is also a halo of positive bias (clouds not reflective enough) around the ocean perimeter of Greenland. The `Arctic` grid has much smaller cloud forcing biases over the Arctic land masses, but is still too reflective over Alaska, the Canadian Archipelago, and parts of Eurasia. Compared to the `Arctic` grid, the `Arctic−GrIS` grid vastly reduces the cloud forcing bias over Eurasia, and also improves the bias over North America. In both VR grids, the halo of positive shortwave cloud forcing bias around the perimeter of Greenland is absent.

The summer cloud forcing biases are consistent with the summer temperature biases in Figure 6 – regions where clouds are too reflective coincide with regions that are too cold. While we have not quantified the role of the clouds in contributing to the cooler Arctic temperatures, shortwave radiation is a crucial component of the Arctic energy budget during summer. The shortwave cloud forcing biases are on the order of 10 W/m2, which is a significant fraction of the total absorbed shortwave during Arctic summer (Serreze et al., 2007) and is therefore likely a factor contributing to the cooler temperatures.

3.2 Clouds and precipitation over Greenland

In addition to summer temperatures, shortwave radiation is an important determinant of snow and ice melt. Figure 8 shows the summer incident shortwave radiation bias at the surface over Greenland and surrounding seas. The top panel shows the bias relative to the RACMO3p2 dataset, and the middle panel relative to the CERES dataset. The halo of excessive incident shortwave radiation around the coasts of Greenland is apparent for both datasets in relation to the coarser grids, consistent with the shortwave cloud forcing biases in Figure 7.

The ice sheet interior receives too little shortwave radiation in the coarser grids. On the VR grids, both the interior shortwave deficit and the excessive shortwave around the ocean perimeter are improved. This suggests that the coarse-grid clouds are too thick in the Greenland interior and too thin around the perimeter. This is consistent with the total summer cloud fraction bias, computed from the CALIPSO cloud dataset and shown in the bottom panel of Figure 8. Note that total cloud fraction characterizes the cloud field at all vertical levels, and attenuates the changes arising from any single layer due to the occurrence of overlapping clouds at other levels. Despite the attenuated signal, the cloud fractions are too high in the interior and and too low around the oceanic perimeter in the coarser grids, whereas the VR grids have an overall improvement in total cloud fraction bias.

The top panel of Figure 9 show the annual climatological mean precipitation bias over the GrIS, expressed as the fractional difference from the RACMO2.3p2 solution. The coarse $1^\circ - 2^\circ$ grids have large, positive biases centered over the southern dome. The `Arctic` grid reduces this bias substantially, and the `Arctic−GrIS` grid reduces it further, with precipitation centers migrating from the interior toward the margins.

The more accurate representation of orographic precipitation in the VR grids is consistent with the cloud and radiation biases, cf. Figures 7, 8, and 9. The agreement of the cloud and precipitation biases in and around Greenland from multiple independent datasets indicates that the biases are a robust feature of the coarser grids. The reduced biases on the VR grids suggest that the biases in the coarse models are a result of insufficient horizontal resolution.

3.3 Greenland surface mass balance

Table 3 shows the 1979-1998 climatological SMB components for each grid, compared with RACMO Andrew - do you want to talk about the values at all?. The CESM values are averages over the ensemble of common ice masks and regridding methods described in section 2.5, and the RACMO values are averages over both RACMO datasets (Table 2) using the same common-ice-mask approach. Table 3 also contains (in parentheses) the SMB components derived from evaluating the integrals on each model's native grid and ice mask. Of note is the large reduction in melt rates using the common-ice-mask approach compared to the native grid, illustrating the dissipation discussed in section 2.5. The errors are greatest in partially ice-covered grid cells straddling the ice

grid name	accumulation	total melt	runoff	sublimation	SMB
RACMO	681.7 (733.5)	-318.6 (-436.4)	-189.1 (-258.5)	-34.5 (-38.8)	458.1 (436.2)
`ne30pg2`	1007. (973.4)	-519.9 (-647.3)	-381.9 (-347.0)	-33.9 (-32.1)	591.2 (594.3)
`ne30pg3`	931.0 (909.3)	-540.8 (-686.7)	-375.8 (-330.1)	-34.1 (-32.6)	521.2 (546.6)
`f19`	884.9 (913.5)	-414.0 (-546.5)	-284.0 (-284.3)	-36.5 (-37.5)	564.4 (591.7)
`f09`	873.9 (882.1)	-389.1 (-482.3)	-256.1 (-212.3)	-37.3 (-37.4)	580.5 (632.4)
`Arctic`	784.1 (818.6)	-335.5 (-436.8)	-215.8 (-194.2)	-42.4 (-43.9)	526.0 (580.5)
`Arctic − GrIS`	693.8 (747.3)	-437.3 (-610.4)	-276.8 (-307.8)	-48.1 (-51.8)	369.0 (387.7)

Table 3. 1979-1998 surface mass balance of the Greenland Ice Sheet in Gt/yr. Values shown are using the common ice mask approach described in the methods section, whereas values in parentheses are from integrating over the native grid and ice mask.

sheet margins, in the ablation zone where melt rates are large. For integrated precipitation, the differences between the native and common-ice-mask approaches are much smaller, since the combined solid/liquid precipitation rates are not directly tied to the ice mask.

Figure 10 shows time series of annually integrated precipitation and snow/ice melt over the GrIS for the various different grids and dycores, with both versions of RACMO shown in black. The 1979-1998 climatological mean values, listed in Table 3, are shown as circles on the right side of the panels. The uniform $1° - 2°$ grids have positive precipitation biases in the interior, whereas the VR grids have the smallest biases, with precipitation comparable to RACMO. The `f19` and `f09` grids perform similarly, with +110 Gt/yr bias, whereas `ne30pg3` is biased by about +165 Gt/yr and `ne30pg2` by +230 Gt/yr. The larger biases on the uniform-resolution SE grids relative to the FV grids are consistent with the coarser GrIS resolution on the SE grids (Figure 3).

The combined annual snow/ice melt shown in the bottom panel of Figure 10 indicates that the `Arctic` grid simulates the most realistic melt rates, with the other grids having more melt than RACMO. The `Arctic−GrIS` grid overpredicts melting by about 125 Gt/yr. This is likely due to an anomalously warm lower troposphere during the summer, relative to the `Arctic` run (Figure 6). The `f19` and `f09` melting rates are improved over `Arctic−GrIS`, overestimating melt by only 70–90 Gt/yr. The SE grids have the largest positive melt bias, between 200–220 Gt/yr. It is more difficult to attribute these differences to resolution alone, since the FV grids have colder summer temperatures than the uniform-resolution SE grids. (Not sure I understand this sentence. FV is both cooler and higher-resolution than SE, so one might suspect that cooler T is, in fact, connected to higher resolution.)

To illustrate the regional behavior of the SMB components, Figure 11 shows the precipitation and combined snow/ice melt integrated over the basins defined by Rignot and Mouginot (2012). The uncertainty due to differences in basin area is larger than for GrIS-wide integrals, owing to the differences in basin boundaries as represented by the common ice masks, which are shown in the `f19` and `ne30pg2` panels of Figure 9. Nonetheless, the regional totals in Figure 11 correctly show the southeast and southwest basins have the most accumulation. In all basins, accumulation decreases monotonically with increasing grid resolution, though with some exceptions. The `Arctic−GrIS` grid simulates less precipitation than RACMO in the central-east and southeast basins, and is closest of all grids to the RACMO precipitation in the large southwest basin.

The basin-integrated melt rates in Figure 11 depend on the dycore. The uniform-resolution SE grids have the largest positive biases in all basins. The `Arctic−GrIS` grid is a close second, while the FV grids have systematically smaller melt-rates. The "second-place" standing of `Arctic−GrIS` is somewhat unexpected, as this grid has the warmest lower-troposphere summer temperatures (Figure 6) and greatest incident shortwave radiation (Figure 8), yet it has less melting than the uniform-resolution SE grids.

Lower troposphere temperature is not a strict proxy for melting; e.g., it may not capture microclimate effects as a result of a better representation of the low-elevation ablation zones. Positive degree-days (PDD; Braithwaite, 1984), which accumulate the near-surface temperature in $^{\circ}C$ for days with temperature above freezing, are a more accurate proxy. PDD is nonlinear in mean monthly temperature (Reeh, 1991). We compute it from monthly mean 2-meter temperature using the method of Calov and Greve (2005), assuming a fixed monthly mean standard deviation of 3°C and a degree-day factor of 5 mm $\mathrm{d}^{-1}\ {}^{\circ}\mathrm{C}^{-1}$.

Figure 11c shows the basin-integrated PDD melt estimate. In the large southeast and southwest basins (and all the other western basins), the `ne30pg3` grid has larger PDD-based melt than the `Arctic−GrIS` grid. The FV grids also have large PDD-based melt in the southwest basin, relative to `Arctic−GrIS`. The PDD plots indicate that the near-surface temperatures (which contribute to melt) are not well approximated by the summer lower-troposphere temperatures in Figure 6.

Figure 9 presents the biases in the combined ice/snow melt as map plots. These plots show that the largest melt biases are on the southeast and northwest coasts, where large coarse-grid cells overlap with the ocean. One possibility is that these problematic grid cells are situated at lower elevations than the true ice sheet surface, leading to a warm bias and too much melt. Figure 12 shows the representation of the ice sheet surface along two transects on the different grids, compared to the high-resolution dataset used to generate CAM topographic boundary conditions (J. Danielson & Gesch, 2011; Lauritzen et al., 2015). The two transects are shown in Figure 9: the east-west “K-transect” in southwest Greenland and a transect extending from the central dome down to the Kangerlussuaq glacier on the southeast coast. The $1^{\circ}-2^{\circ}$ grids are noticeably coarse, with only a few grid cells populating the transect. The `f09` grid is a bit of an exception for the K-transect, with grid cells becoming narrow in the meridional direction at high latitudes. The VR grids are more skillful at reproducing the steep margins of the ice sheet, capturing the parabolic shape of the GrIS margins.

The transects in Figure 12 show that the ice sheet surface on the coarse grids is not systematically lower than the true surface in ablation zones. Rather, the smoothing and flattening of the raw topography, necessary to prevent the model from exciting grid-scale numerical modes, causes the lower-elevation ablation zones to extend beyond the true ice sheet margin, where they lie above the actual ice surface. The `f19` grid has both the smoothest topography and the flattest ice sheet since its dynamics are coarsest (`f09`, `ne30pg2` and `ne30pg3` use identical smoothing). This suggests that coarser models will tend to elevate the ablation zones and thereby depress melt rates.

Figure 12 also shows the ice margin boundary, illustrating that the ablation zone lies in a narrow horizontal band where the ice sheet rapidly plunges to sea-level. Due to this abrupt transition, coarse grids will commonly represent the ablation zone with grid cells containing mixtures of ice-covered and ice-free regions. We hypothesize that coarser models have larger melt biases because summer melting is confined to these mixed ice/land/ocean grid cells. CLM deals with land heterogeneity in a complex and sophisticated manner, but CAM only sees a homogenized state due to volume averaging over the sub-grid mixture. Thus, warm ice-free land patches in a grid cell may unduly influence the climate over the entire grid cell, causing a warm bias over the ice-covered patch. (This is an interesting conclusion pointing to the need for better treatment of surface inhomogeneity in CAM. This might be a way to compensate for coarse resolution in future CESM versions?)

Figure 13 shows mean melt bias, relative to both RACMO datasets, conditionally sampled based on grid cell ice fraction in the GrIS region. Errors are computed using the common-ice-mask approach, meaning that all fields are mapped to the common masks, which define the grid cell ice fraction. The figure shows (Any idea why the errors are smaller in the cells with intermediate ice fraction? I wouldn't have expected this.) that coarser grids generally have two peaks in ice fraction space; a bump in positive melting errors in the 0-20% range, and another in the fully ice-covered cells. Also shown are the ± 1 standard deviation of the biases for each bin. They indicate that the biases in 0-20% bins are mostly contained in the positive bias region (a fractional bias greater than 0), whereas the fully-covered ice cells have a wider distribution, with many grid cells also containing negative melting biases. The excessive melting in the 0-20% ice fraction bins supports our hypothesis that the prevalence of mixed-grid cells in the ablation zone on coarse grids is responsible for their large melt bias.

Rene - is the melt map consistent with the bin figure? Smallest errors are interior points which correspond to glacier fraction bin of 1, which the largest errors. ARH - I can check. The bin figure is fractional change so its a different metric, which could explain the apparent inconsistency.

3.4 Precipitation extremes

Synoptic storms are tracked using TempestExtremes atmospheric feature detection software (Ullrich et al., 2021). As the `Arctic` grid contains 1/4° refinement north of about 45° latitude, the storm tracker is applied to this region for the `Arctic` and `ne30pg3` runs to identify differences in storm characteristics due to horizontal resolution. The composite mean precipitation maps are similar between the two grids, and exhibit the iconic comma structure of synoptic cyclones (not shown). Marcus - no need to mention if not shown.

Figure 14 shows monthly PDFs of the precipitation rates associated with storms. The PDFs are constructed by sampling all the precipitation rates within 30° of the storm center, for each point on the storm track and for all storms. The PDFs are evaluated on an identical composite grid for all runs, and so storm statistics are not impacted by differences in output resolution. The `Arctic` run has larger extreme precipitation rates compared to `ne30pg3` in every month, but the increase is greatest in the summer months, which coincides with the most extreme events of the year. This is primarily due to increased resolution and not the reduced physics times-step; the `ne30pg3`* run only marginally increases the extreme precipitation rates compared with `ne30pg3` (Figure 14).

The extreme precipitation rates in the `Arctic` run are closer than `ne30pg3` to the ERA5 reanalysis (Figure 14). It is difficult to know how much the extreme precipitation rates in ERA5 are constrained by data assimilation, or whether these precipitation rates are due to using a similar 1/4° model as the `Arctic` grid. However, it is well documented that 1/4° models are more skillful at simulating extreme events (Bacmeister et al., 2013; Obrien et al., 2016). A more realistic representation of extreme precipitation events is an additional benefit of the VR grids.

4 Conclusions

Running CESM2.2 in an AMIP-style configuration, we have evaluated six grids from two dynamical cores for their performance over the Arctic and in simulating the GrIS SMB. The $1 - 2°$ finite-volume grids have enhanced resolution over polar regions due to the convergence of meridian lines, although a polar filter is used to prevent spurious atmospheric features from forming at this higher resolution. Spectral-element grids comparable to the resolution of the finite-volume grids have an isotropic grid structure where the grid resolution is similar over the entire model domain. We developed two VR grids and introduced them into CESM2.2 as part of this work. Both use the spectral-element dycore; the `Arctic` grid has $1/4°$ refinement over the broader Arctic, whereas the `Arctic−GrIS` grid is identical except for a $1/8°$ patch of refinement over Greenland. A third VR grid, `CONUS`, is also available in CESM2.2.

In general, the FV grids have colder summer temperatures over the Arctic compared with the SE grids (including the VR grids). The cloud biases in all the uniform-resolution grids, whether FV or SE, are similar, in general being too cloudy over Arctic land masses.Marcus - hard to parse this sentence. The VR grids reduce the cloud biases. It should be emphasized that our analysis is specific to the Arctic summer because of its relevance to GrIS melt rates; an improved representation of clouds in the Arctic does not imply improved clouds at lower latitudes.

At the regional level, there is a halo of negative cloud bias (/colorblueI got confused about signs here, because section 3 talks about a halo of positive cloud SW forcing bias, which corresponds to a negative bias in cloud amounts. Maybe replace 'cloud bias' with 'cloudiness bias' or something similar?) around the ocean perimeter of Greenland on all $1{-}2°$ grids, but not the VR grids. This halo bias coincides with a positive cloud bias over the ice sheet interior. This anomaly pattern has been attributed to deficient orographic precipitation on the coarser model grids. With overly smooth topography on the $1{-}2°$ grids, synoptic systems moving into Greenland are not sufficiently lifted when encountering the steep ice margins. As a result, excess precipitation falls in the GrIS interior, instead of being concentrated on the steep coastal margins as shown by observations. This results in a positive precipitation and cloud bias in the ice sheet interior, and a halo of low cloud bias about the perimeter. The agreement of different observational data products on this bias lends confidence in the attribution of causes. The VR grids compare better to the observations and show that orographic precipitation in Greenland is largely resolved when the horizontal resolution is increased sufficiently.

We integrated the primary source and sink terms of the SMB equation over the GrIS for each of the six grids. The uniform $1{-}2°$ grids have large positive accumulation biases because they fail to resolve orographic precipitation. The uniform SE grids have larger accumulation biases, suggesting that the FV grids are more skillful for precipitation due to finer resolution over Greenland, despite a polar filter. The VR grids have the most accurate accumulation rates of all the grids.

The primary mass sink term of the GrIS, ice/snow melt, has similar biases. The uniform resolution SE grids have too much melt, while the FV grids have smaller biases. It is difficult to attribute these biases to grid resolution alone. The FV grids have colder summers, consistent with their lower melt bias. However, the `Arctic − GrIS` grid has the warmest summer temperatures of all grids, yet it has less melting than the uniform-resolution SE grids. This suggests that grid resolution is responsible for a large fraction of the melt biases. We propose a mechanism: Coarse grids represent ablation zones using grid cells with mixed surface types, ice-covered and ice-free. The warmer ice-free patches may largely determine the mean state, leading to a warm bias over the ice-covered patches of the grid cell. This mechanism is supported by analysis of melt biases binned by grid-cell ice fraction.

The `Arctic` grid substantially improves the simulated Arctic climate, including precipitation extremes and the Greenland SMB, compared to the uniform $1° - 2°$ grids. The `Arctic−GrIS` grid has the most realistic cloud and precipitation fields, but its summer temperatures are too warm. The $1°$ FV model gives a surprisingly realistic SMB, likely due to the relatively fine resolution of Greenland on lat-lon grids. It is also the most heavily tuned model configuration. May be worth mentioning here as well. In particular, a greater number of grid cells in the ablation zone reduces the influence of mixed ice-covered/ice-free grid cells that represent ablation poorly on the other uniform-resolution grids.

As modeling systems move away from lat-lon grids towards quasi-uniform unstructured grids, it is worth taking stock of whether this will degrade the simulated polar climate. We have found that the 1° FV model has clear advantages over the 1° SE model in simulating the surface mass balance of the GrIS. This finding will not interrupt the ongoing transition towards unstructured grids in CESM, largely driven by gains in computational efficiency, but it has inspired us to develop alternative configurations that recover or improve on the fidelity of polar climate. We have shown here (and in a prior companion study (van Kampenhout et al., 2018)) that for CESM, Arctic-refined meshes can substantially improve the simulated mass balances of the GrIS, even compared to the 1° grid. This should reassure the CESM modeling community that the ongoing transition away from lat-lon grids will not adversely impact CESM's usefuless as a state-of-the-art tool for simulating and understanding polar processes. (WHL: This last sentence may be too sanguine. Yes, we can recover the fidelity of Arctic simulations using VR grids, but (so far) only at a considerable cost in cpu-hours. This points to the need for an intermediate resolution that is more affordable, and/or model development or tuning that reduces the biases on coarse grids.)Andrew - Maybe better to just state that higher resolution is better: 1deg better than 2deg FV, and ne30 is coarser than 1deg at Greenland latitudes so it's worse. But VR has higher resolution so SR-VR is better. It's all about resolution.

We are working to develop a configuration of the `Arctic` grid that is fully-coupled with the CESM ocean and sea ice components and the Community Ice Sheet Model (CISM), to provide multi-century projections of the state of the GrIS and its contribution to sea-level rise. We have also developed a visualization of the `Arctic−GrIS` run, now available on youtube[1]. Figure 15 shows a snapshot of this visualization, illustrating mesoscale katabatic winds descending the southeastern slopes of GrIS. These new grids and configurations will provide new opportunities for CESM polar science and aims to contribute to an improved understanding of the polar environment. (WHL: I replaced the previous last sentence because it seemed too much like an advertisement. However, this new ending seems weak. I wonder if we should say something about future work motivated by this study, for instance investigating grids and parameterizations that provide some of the same benefits as these VR grids but at lower cost.)

Appendix A Details on spectra-element dynamical core improvements since the CESM2.0 release

Since the CESM2.0 release of the spectral-element dynamical core documented in Lauritzen et al. (2018) some important algorithmic improvements have been implemented and released with CESM2.2. These pertain mainly to the flow over orography that, for the spectral-element dynamical core, can lead to noise aligned with the element boundaries (Herrington et al., 2018).

A1 Reference profiles

Significant improvement in removing noise for flow over orography can be achieved by using reference profiles for temperature and pressure

$$T^{(ref)} = T_0 + T_1 \Pi^{(ref)}, \tag{A1}$$
$$p_s^{(ref)} = p_0 \exp\left(-\frac{\Phi_s}{R^{(d)} T_{ref}}\right), \tag{A2}$$

(Simmons & Jiabin, 1991) where g gravity, $T_1 = \Gamma_0 T_{ref} c_p^{(d)}/g \approx 192K$ with standard lapse rate $\Gamma_0 \equiv 6.5K/km$ and $T_0 \equiv T_{ref} - T_1 \approx 97K$; $T_{ref} = 288K$ ($c_p^{(d)}$ specific heat

[1] `https://www.youtube.com/watch?v=YwHgqDu75s8&t=4s&ab_channel=NCARVisLab`

of dry air at constant pressure; $R^{(d)}$ gas constant for dry air), and Φ_s surface geopotential. The reference Exner function is

$$\Pi^{(ref)} = \left(\frac{p^{(ref)}}{p_0}\right)^{\kappa} \tag{A3}$$

where $\kappa = \frac{R^{(d)}}{c_p^{(d)}}$. The reference surface pressure $p_0 = 1000hPa$ and at each model level the reference pressure $p^{(ref)}$ is computed from $p_s^{(ref)}$ and the standard hybrid coefficients

$$p^{(ref)}(\eta) = A(\eta)p_0 + B(\eta)p_s^{(ref)}, \tag{A4}$$

where A and B are the standard hydrid coefficients (using a dry-mass generalized vertical mass coordinate η). These reference profiles are subtracted from the prognostic temperature and pressure-level-thickness states before applying hyperviscosity:

$$\text{CESM2.0} \quad \rightarrow \quad \text{CESM2.2}$$

$$\nabla^4_\eta T \quad \rightarrow \quad \nabla^4_\eta \left(T - T^{(ref)}\right), \tag{A5}$$

$$\nabla^4_\eta \delta p^{(d)} \quad \rightarrow \quad \nabla^4_\eta \left(\delta p^{(d)} - \delta p^{(ref)}\right). \tag{A6}$$

This reduces spurious transport of temperature and mass up/down-slope due to the hyperviscosity operator.

A2 Rewriting the pressure gradient force (PGF)

In the CESM2.0 the following (standard) form of the pressure gradient term was used:

$$\nabla_\eta \Phi + \frac{1}{\rho}\nabla_\eta p, \tag{A7}$$

where Φ is geopotential and $\rho = \frac{R^{(d)}T_v}{p}$ is density (for details see Lauritzen et al., 2018). To alleviate noise for flow over orography, we switched to an Exner pressure formulation following Taylor et al. (2020), which uses that (A7) can be written in terms of the Exner pressure

$$\nabla_\eta \Phi + c_p^{(d)}\theta_v \nabla_\eta \Pi, \tag{A8}$$

where the Exner pressure is

$$\Pi \equiv \left(\frac{p}{p_0}\right)^{\kappa}. \tag{A9}$$

The derivation showing that (A7) and (A8) are equivalent is shown here:

$$\begin{aligned}
c_p^{(d)}\theta_v \nabla_\eta \Pi &= c_p^{(d)}\theta_v \nabla_\eta \left(\frac{p}{p_0}\right)^{\kappa}, \\
&= c_p^{(d)}\theta_v \kappa \left(\frac{p}{p_0}\right)^{\kappa-1} \nabla_\eta \left(\frac{p}{p_0}\right), \\
&= c_p^{(d)}\theta_v \kappa \Pi \left(\frac{p_0}{p}\right) \nabla_\eta \left(\frac{p}{p_0}\right), \\
&= \frac{c_p^{(d)}\theta_v \kappa \Pi}{p}\nabla_\eta p, \\
&= \frac{R^{(d)}\theta_v \Pi}{p}\nabla_\eta p, \\
&= \frac{R^{(d)}T_v}{p}\nabla_\eta p, \\
&= \frac{1}{\rho}\nabla_\eta p.
\end{aligned}$$

Using the reference states from (Simmons & Jiabin, 1991),

$$\overline{T} = T_0 + T_1\Pi, \tag{A10}$$
$$\overline{\theta} = T_0/\Pi + T_1, \tag{A11}$$

we can define a geopotential as a function of Exner pressure

$$\overline{\Phi} = -c_p^{(d)}\left(T_0 \log \Pi + T_1\Pi - T_1\right). \tag{A12}$$

This "balanced" geopotential obeys

$$c_p^{(d)}\overline{\theta}\nabla\Pi + \nabla\overline{\Phi} = 0 \tag{A13}$$

for any Exner pressure. Subtracting this "reference" profile from the PGF yields

$$\begin{aligned}\nabla_\eta\Phi + c_p^{(d)}\theta_v\nabla_\eta\Pi &= \nabla_\eta\left(\Phi - \overline{\Phi}\right) + c_p^{(d)}\left(\theta_v - \overline{\theta}\right)\nabla_\eta\Pi,\\ &= \nabla_\eta\Phi + c_p^{(d)}\theta_v\nabla_\eta\Pi + c_p^{(d)}T_0\left[\nabla_\eta \log\Pi - \frac{1}{\Pi}\nabla_\eta\Pi\right].\end{aligned} \tag{A14}$$

In the continuum, the two formulations (left and right-hand side of (A14)) are identical. But under discretization, the second formulation can have much less truncation error.

A3 Results

[Adam: have you defined ne30np4 in the main text?]

One year averages of vertical pressure velocity at 500hPa (`OMEGA500`) have been found to be a useful quantity to detect spurious up or down-drafts induced by steep orography (Figure A1). While the true solution is not known, strong vertical velocities aligned with element edges that are not found in the CAM-FV reference solution (Figure A1(a)) are likely not physical (spurious). The older CESM2.0 version of SE (Figure A1(d)) using the "traditional" discretization of the PGF, (A14), exhibits significant spurious noise patters around steep orography compared to CAM-FV (e.g., around Himalayas and Andes). This is strongly alleviated by switching to the Exner formulation of the PGF (A8; Figure A1(c)). By also subtracting reference profiles from pressure-level thickness and temperature, equations (A5) and (A6) respectively, reduces strong up-down drafts further (Figure A1(d)). Switching to the CAM-SE-CSLAM version where physics tendencies are computed on an quasi-equal area physics grid and using the CSLAM transport scheme, marginal improvements are observed in terms of a smoother vertical velocity field (Figure A1(e,f)). The configuration shown in Figure A1(d) is used for the simulations shown in the main text of this paper.

It is interesting to note that the noise issues and algorithmic remedies found in the real-world simulations discussed above, can be investigated by replacing all of physics with a modified version of the Held-Suarez forcing (Held & Suarez, 1994). The original formulation of the Held-Suarez idealized test case used a flat Earth ($\Phi_s = 0$) and a dry atmosphere. By simply adding the surface topography used in 'real-world' simulations and removing the temperature relaxation in the lower part of domain ($\sigma > 0.7$; see Held and Suarez (1994) for details), surprisingly realistic vertical velocity fields (in terms of structure) result (see Figure A2). Since this was a very useful development tool it is shared in this manuscript.

Acknowledgments

This material is based upon work supported by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the NSF under Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. A. Herrington thanks Matt Rehme (NCAR/CISL) for his role in generating the `Arctic−GrIS` visualization available on youtube (`https://www.youtube.com/watch?v=YwHgqDu75s8&t=4s&ab_channel=NCARVisLab`).

The data presented in main part of this manuscript is available at `https://github.com/adamrher/2020-arcticgrids`. The source code and data for the Appendix is available at `https://github.com/PeterHjortLauritzen/CAM/tree/topo-mods`.

References

Bacmeister, J. T., & Coauthors. (2018). Projected changes in tropical cyclone activity under future warming scenarios using a high-resolution climate model. *Climatic Change*, *146*, 547-560. Retrieved from `http://dx.doi.org/10.1007/s10584-016-1750-x` doi: 10.1007/s10584-016-1750-x

Bacmeister, J. T., Wehner, M. F., Neale, R. B., Gettelman, A., Hannay, C., Lauritzen, P. H., ... Truesdale, J. E. (2013). Exploratory high-resolution climate simulations using the community atmosphere model (cam). *J. Climate*, *27*(9), 3073–3099. doi: 10.1175/JCLI-D-13-00387.1

Bambach, N. E., Rhoades, A. M., Hatchett, B. J., Jones, A. D., Ullrich, P. A., & Zarzycki, C. M. (2021). Projecting climate change in south america using variable-resolution community earth system model: An application to chile. *International Journal of Climatology*.

Beljaars, A., Brown, A., & Wood, N. (2004). A new parametrization of turbulent orographic form drag. *Quart. J. Roy. Meteor. Soc.*, *130*(599), 1327–1347. doi: 10.1256/qj.03.73

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*(23), 9655–9676.

Box, J. E., Bromwich, D. H., & Bai, L.-S. (2004). Greenland ice sheet surface mass balance 1991–2000: Application of polar mm5 mesoscale model and in situ data. *Journal of Geophysical Research: Atmospheres*, *109*(D16).

Braithwaite, R. J. (1984). Calculation of degree-days for glacier-climate research. *Zeitschrift für Gletscherkunde und Glazialgeologie*, *20*(1984), 1–8.

Burakowski, E. A., Tawfik, A., Ouimette, A., Lepine, L., Zarzycki, C., Novick, K., ... Bonan, G. (2019). Simulating surface energy fluxes using the variable-resolution community earth system model (vr-cesm). *Theoretical and Applied Climatology*, *138*(1), 115–133.

Calov, R., & Greve, R. (2005). A semi-analytical solution for the positive degree-day model with stochastic temperature variations. *Journal of Glaciology*, *51*(172), 173–175.

Canuto, C., Hussaini, M. Y., Quarteroni, A., & Zang, T. (2007). *Spectral methods: Evolution to complex geometries and applications to fluid dynamics* (1st ed.). Springer.

Chang, P., Zhang, S., Danabasoglu, G., Yeager, S. G., Fu, H., Wang, H., ... others (2020). An unprecedented set of high-resolution earth system simulations for understanding multiscale interactions in climate variability and change. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002298.

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

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., ... Zhang, M. (2006). The formulation and atmospheric simulation of the community atmosphere model version 3 (cam3). *Journal of Climate*, *19*(11), 2144–2161.

Copernicus, C. (2019). Era5 monthly averaged data on pressure levels from

1979 to present. *URL: https://cds. climate. copernicus. eu*. doi: 10.24381/cds.6860a573

Craig, C., Bacmeister, J., Callaghan, P., Eaton, B., Gettelman, A., Goldhaber, S. N., … Vitt, F. M. (2021). *Cam6.3 user's guide* (Tech. Rep.). NCAR/TN-571+EDD. doi: 10.5065/Z953-ZC95

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., … others (2020). The community earth system model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916.

Danielson, J., & Gesch, D. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)* (Open-File Report 2011-1073). U.S. Geological Survey. (http://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf)

Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)* (Open File Rep. No. 2011-1073). US Geological Survey. doi: https://doi.org/10.3133/ofr20111073

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., … Worley, P. H. (2012). CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *Int. J. High. Perform. C.*, *26*(1), 74-89. Retrieved from `http://hpc.sagepub.com/content/26/1/74.abstract` doi: 10.1177/1094342011428142

Evans, K. J., Kennedy, J. H., Lu, D., Forrester, M. M., Price, S., Fyke, J., … others (2019). Livvkit 2.1: automated and extensible ice sheet model validation. *Geoscientific Model Development*, *12*(3), 1067–1086.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Fettweis, X., Franco, B., Tedesco, M., Van Angelen, J., Lenaerts, J. T., van den Broeke, M. R., & Gallée, H. (2013). Estimating the greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model mar. *The Cryosphere*, *7*(2), 469–489.

Gettelman, A., Callaghan, P., Larson, V., Zarzycki, C., Bacmeister, J., Lauritzen, P., … Neale, R. (2017). Regional climate simulations with the community earth system model. *J. Adv. Model. Earth Syst.*. (submitted)

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A., Danabasoglu, G., … others (2019). High climate sensitivity in the community earth system model version 2 (cesm2). *Geophysical Research Letters*, *46*(14), 8329–8337.

Gettelman, 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*(3), 1268–1287.

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., & Caldwell, P. (2015). Advanced two-moment bulk microphysics for global models. part ii: Global model solutions and aerosol–cloud interactions. *Journal of Climate*, *28*(3), 1288–1307.

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A pdf-based model for boundary layer clouds. part i: Method and model description. *Journal of the Atmospheric Sciences*, *59*(24), 3540–3551. doi: 10.1175/1520-0469(2002)059⟨3540:apbmfb⟩2.0.co;2

Guba, O., Taylor, M. A., Ullrich, P. A., Overfelt, J. R., & Levy, M. N. (2014). The spectral element method (sem) on variable-resolution grids: evaluating grid sensitivity and resolution-aware numerical viscosity. *Geosci. Model Dev.*, *7*(6), 2803–2816. doi: 10.5194/gmd-7-2803-2014

Guo, Z., Wang, M., Qian, Y., Larson, V. E., Ghan, S., Ovchinnikov, M., … Zhou, T. (2015). Parametric behaviors of clubb in simulations of low clouds in the c ommunity a tmosphere m odel (cam). *Journal of Advances in Modeling Earth Systems*, *7*(3), 1005–1025.

Hansen, N., Simonsen, S. B., Boberg, F., Kittel, C., Orr, A., Souverijns, N., ... Mottram, R. (2022). Brief communication: Impact of common ice mask in surface mass balance estimates over the antarctic ice sheet. *The Cryosphere*, *16*(2), 711–718.

Held, I. M., & Suarez, M. J. (1994). A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Am. Meteorol. Soc.*, *73*, 1825-1830.

Herrington, A. R., Lauritzen, P., Taylor, M. A., Goldhaber, S., Eaton, B. E., Bacmeister, J., ... Ullrich, P. (2018). Physics-dynamics coupling with element-based high-order galerkin methods: quasi equal-area physics grid. *Mon. Wea. Rev.*, *47*, 69–84. doi: 10.1175/MWR-D-18-0136.1

Herrington, A. R., Lauritzen, P. H., Reed, K. A., Goldhaber, S., & Eaton, B. E. (2019). Exploring a lower resolution physics grid in cam-se-cslam. *Journal of Advances in Modeling Earth Systems*, *11*.

Herrington, A. R., & Reed, K. A. (2018). An idealized test of the response of the community atmosphere model to near-grid-scale forcing across hydrostatic resolutions. *J. Adv. Model. Earth Syst.*, *10*(2), 560–575.

Herrington, A. R., & Reed, K. A. (2020). On resolution sensitivity in the community atmosphere model. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 3789–3807.

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *Journal of Climate*, *21*(19), 5145–5153.

Jablonowski, C., & Williamson, D. L. (2011). The pros and cons of diffusion, filters and fixers in atmospheric general circulation models. In P. H. Lauritzen, C. Jablonowski, M. Taylor, & R. Nair (Eds.), *Numerical techniques for global atmospheric models* (pp. 381–493). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-11640-7_13

Lauritzen, P. H., Bacmeister, J. T., Callaghan, P. F., & Taylor, M. A. (2015). Ncar global model topography generation software for unstructured grids. *Geoscientific Model Development Discussions*, *8*(6), 4623–4651. doi: 10.5194/gmdd-8-4623-2015

Lauritzen, P. H., Jablonowski, C., Taylor, M., & Nair, R. D. (2010). Rotated versions of the jablonowski steady-state and baroclinic wave test cases: A dynamical core intercomparison. *J. Adv. Model. Earth Syst.*, *2*(15), 34 pp.

Lauritzen, P. H., Mirin, A., Truesdale, J., Raeder, K., Anderson, J., Bacmeister, J., & Neale, R. B. (2011). Implementation of new diffusion/filtering operators in the CAM-FV dynamical core. *Int. J. High Perform. Comput. Appl.*. doi: 10.1177/1094342011410088

Lauritzen, P. H., Nair, R., Herrington, A., Callaghan, P., Goldhaber, S., Dennis, J., ... Dubos, T. (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. *J. Adv. Model. Earth Syst.*, *10*(7), 1537-1570. doi: 10.1029/2017MS001257

Lauritzen, P. H., Taylor, M. A., Overfelt, J., Ullrich, P. A., Nair, R. D., Goldhaber, S., & Kelly, R. (2017). CAM-SE-CSLAM: Consistent coupling of a conservative semi-lagrangian finite-volume method with spectral element dynamics. *Mon. Wea. Rev.*, *145*(3), 833-855. doi: 10.1175/MWR-D-16-0258.1

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., ... others (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*(12), 4245–4287.

Lin, S.-J. (2004). A 'vertically Lagrangian' finite-volume dynamical core for global models. *Mon. Wea. Rev.*, *132*, 2293-2307.

Lin, S.-J., & Rood, R. B. (1997). An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Q.J.R.Meteorol.Soc.*, *123*, 2477-2498.

Lipscomb, W. H., Fyke, J. G., Vizcaíno, M., Sacks, W. J., Wolfe, J., Vertenstein, M., … Lawrence, D. M. (2013). Implementation and initial evaluation of the glimmer community ice sheet model in the community earth system model. *Journal of Climate*, *26*(19), 7352–7371.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., … Kato, S. (2018). Clouds and the earth's radiant energy system (ceres) energy balanced and filled (ebaf) top-of-atmosphere (toa) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918.

Lofverstrom, M., Fyke, J. G., Thayer-Calder, K., Muntjewerf, L., Vizcaino, M., Sacks, W. J., … Bradley, S. L. (2020). An efficient ice sheet/earth system model spin-up procedure for cesm2-cism2: Description, evaluation, and broader applicability. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS001984.

Mottram, R., Boberg, F., Langen, P., Yang, S., Rodehacke, C., Christensen, J. H., & Madsen, M. S. (2017). Surface mass balance of the greenland ice sheet in the regional climate model hirham5: Present state and future prospects. *Low Temp. Sci.*, *75*, 105–115.

Muntjewerf, L., Sacks, W. J., Lofverstrom, M., Fyke, J., Lipscomb, W. H., Ernani da Silva, C., … Sellevold, R. (2021). Description and Demonstration of the Coupled Community Earth System Model v2–Community Ice Sheet Model v2 (CESM2-CISM2). *Journal of Advances in Modeling Earth Systems*, *13*(6), e2020MS002356.

Neale, R. B., Richter, J. H., & Jochum, M. (2008). The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate*, *21*, 5904-5924.

Noël, B., Van De Berg, W., Van Meijgaard, E., Kuipers Munneke, P., Van De Wal, R., & Van Den Broeke, M. (2015). Evaluation of the updated regional climate model racmo2. 3: summer snowfall impact on the greenland ice sheet. *The Cryosphere*, *9*(5), 1831–1844.

Noël, B., van de Berg, W. J., Lhermitte, S., & van den Broeke, M. R. (2019). Rapid ablation zone expansion amplifies north greenland mass loss. *Science advances*, *5*(9), eaaw0123.

Noël, B., van de Berg, W. J., Van Wessem, J. M., Van Meijgaard, E., Van As, D., Lenaerts, J., … others (2018). Modelling the climate and surface mass balance of polar ice sheets using racmo2–part 1: Greenland (1958–2016). *The Cryosphere*, *12*(3), 811–831.

Obrien, T. A., Collins, W. D., Kashinath, K., Rübel, O., Byna, S., Gu, J., … Ullrich, P. A. (2016). Resolution dependence of precipitation statistical fidelity in hindcast simulations. *J. Adv. Model. Earth Syst.*, *8*(2), 976–990. Retrieved from `http://dx.doi.org/10.1002/2016ms000671` doi: 10.1002/2016ms000671

Ohmura, A. (2001). Physical basis for the temperature-based melt-index method. *Journal of applied Meteorology*, *40*(4), 753–761.

Pfister, G. G., Eastham, S. D., Arellano, A. F., Aumont, B., Barsanti, K. C., Barth, M. C., … others (2020). The multi-scale infrastructure for chemistry and aerosols (musica). *Bulletin of the American Meteorological Society*, *101*(10), E1743–E1760.

Pollard, D. (2010). A retrospective look at coupled ice sheet–climate modeling. *Climatic Change*, *100*(1), 173–194.

Pollard, D., & Groups, P. P. (2000). Comparisons of ice-sheet surface mass budgets from paleoclimate modeling intercomparison project (pmip) simulations. *Global and Planetary Change*, *24*(2), 79–106.

Pope, V., & Stratton, R. (2002). The processes governing horizontal resolution sensitivity in a climate model. *Climate Dynamics*, *19*(3-4), 211–236.

Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, *227*(1), 55-78.

Rae, J., Adalgeirsdóttir, G., Edwards, T. L., Fettweis, X., Gregory, J., Hewitt, H., ... others (2012). Greenland ice sheet surface mass balance: evaluating simulations and making projections with regional climate models. *The Cryosphere*, *6*(6), 1275–1294.

Rasch, P. J., & Williamson, D. L. (1990). Computational aspects of moisture transport in global models of the atmosphere. *Q. J. R. Meteorol. Soc.*, *116*, 1071-1090.

Reeh, N. (1991). Parameterization of melt rate and surface temperature in the greenland ice sheet. *Polarforschung*, *59*(3), 113–128.

Rhoades, A. M., Huang, X., Ullrich, P. A., & Zarzycki, C. M. (2016). Characterizing sierra nevada snowpack using variable-resolution cesm. *Journal of Applied Meteorology and Climatology*, *55*(1), 173-196. Retrieved from `https://doi.org/10.1175/JAMC-D-15-0156.1` doi: 10.1175/JAMC-D-15-0156.1

Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity wave source parameterization in a general circulation model. *J. Atmos. Sci.*, *67*, 136-156. doi: dx.doi.org/10.1175/2009JAS3112.1

Rignot, E., & Mouginot, J. (2012). Ice flow in greenland for the international polar year 2008–2009. *Geophysical Research Letters*, *39*(11).

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., ... Schulzweida, U. (2006). Sensitivity of simulated climate to horizontal and vertical resolution in the echam5 atmosphere model. *Journal of Climate*, *19*(16), 3771–3791.

Sellevold, R., Van Kampenhout, L., Lenaerts, J., Noël, B., Lipscomb, W. H., & Vizcaino, M. (2019). Surface mass balance downscaling through elevation classes in an earth system model: Application to the greenland ice sheet. *The Cryosphere*, *13*(12), 3193–3208.

Serreze, M. C., Barrett, A. P., Slater, A. G., Steele, M., Zhang, J., & Trenberth, K. E. (2007). The large-scale energy budget of the arctic. *Journal of Geophysical Research: Atmospheres*, *112*(D11).

Simmons, A. J., & Jiabin, C. (1991). The calculation of geopotential and the pressure gradient in the ECMWF atmospheric model: Influence on the simulation of the polar atmosphere and on temperature analyses. *Quart. J. Roy. Meteor. Soc.*, *117*(497), 29-58. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711749703` doi: https://doi.org/10.1002/qj.49711749703

Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., ... Vertenstein, M. (2014). A new synoptic scale resolving global climate simulation using the community earth system model. *J. Adv. Model. Earth Syst.*, *6*(4), 1065–1094. doi: 10.1002/2014MS000363

Suarez, M. J., & Takacs, L. L. (1995). Volume 5 documentation of the aries/geos dynamical core: Version 2.

Taylor, M. A., & Fournier, A. (2010). A compatible and conservative spectral element method on unstructured grids. *J. Comput. Phys.*, *229*(17), 5879 - 5895. doi: 10.1016/j.jcp.2010.04.008

Taylor, M. A., Guba, O., Steyer, A., Ullrich, P. A., Hall, D. M., & Eldred, C. (2020). An energy consistent discretization of the nonhydrostatic equations in primitive variables. *Journal of Advances in Modeling Earth Systems*, *12*(1). doi: 10.1029/2019MS001783

Taylor, M. A., Tribbia, J., & Iskandarani, M. (1997). The spectral element method for the shallow water equations on the sphere. *J. Comput. Phys.*, *130*, 92-108.

Team, E. J. S., Balaji, V., Boville, B., Collins, N., Craig, T., Cruz, C., ... others

(2021). *Esmf user guide* (Tech. Rep.).

Ullrich, P. A., & Taylor, M. A. (2015). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part i. *Monthly Weather Review*, *143*(6), 2419–2440.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). Tempestextremes v2. 1: a community framework for feature detection, tracking and analysis in large datasets. *Geoscientific Model Development Discussions*, 1–37.

Van Angelen, J., Lenaerts, J., Lhermitte, S., Fettweis, X., Kuipers Munneke, P., Van den Broeke, M., … Smeets, C. (2012). Sensitivity of greenland ice sheet surface mass balance to surface albedo parameterization: a study with a regional climate model. *The Cryosphere*, *6*(5), 1175–1186.

van Kampenhout, L., Lenaerts, J. T., Lipscomb, W. H., Lhermitte, S., Noël, B., Vizcaíno, M., … van den Broeke, M. R. (2020). Present-day greenland ice sheet climate and surface mass balance in cesm2. *Journal of Geophysical Research: Earth Surface*, *125*(2).

van Kampenhout, L., Rhoades, A. M., Herrington, A. R., Zarzycki, C. M., Lenaerts, J. T. M., Sacks, W. J., & van den Broeke, M. R. (2018). Regional grid refinement in an earth system model: Impacts on the simulated greenland surface mass balance. *The Cryosphere Discuss.*. doi: 10.5194/tc-2018-257

Wan, H., Giorgetta, M. A., Zängl, G., Restelli, M., Majewski, D., Bonaventura, L., … others (2013). "the icon-1.2 hydrostatic atmospheric dynamical core on triangular grids, part i: formulation and performance of the baseline version". *Geosci. Model Dev.*, *6*, 735–763.

Williamson, D. (2007). The evolution of dynamical cores for global atmospheric models. *J. Meteor. Soc. Japan*, *85*, 241-269.

Williamson, D. (2008). Convergence of aqua-planet simulations with increasing resolution in the community atmospheric model, version 3. *Tellus A*, *60*(5), 848–862. doi: 10.1111/j.1600-0870.2008.00339.x

Zarzycki, C. M., Jablonowski, C., & Taylor, M. A. (2014). Using variable-resolution meshes to model tropical cyclones in the community atmosphere model. *Mon. Wea. Rev.*, *142*(3), 1221-1239. doi: 10.1175/MWR-D-13-00179.1

Zhang, G., & McFarlane, N. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate center general-circulation model. *ATMOSPHERE-OCEAN*, *33*(3), 407-446.

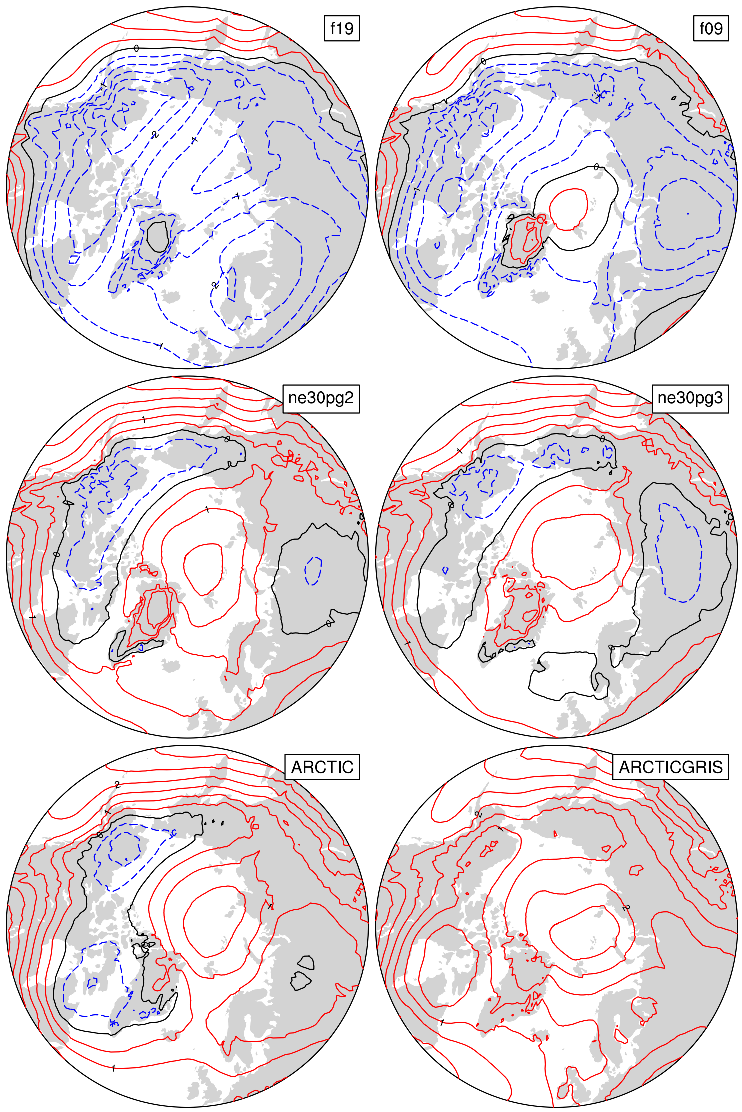

Figure 6. 1979-1998 lower troposphere, northern hemisphere summer virtual temperature biases, computed as the difference from ERA5. Lower troposphere layer mean virtual temperature is derived from the 1000 hPa - 500h Pa geopotential thickness, using the hypsometric equation. Differences are computed after mapping the ERA5 data to the finite-volume grids since the geopotential field is only available on the output tapes in the spectral-element runs that have been interpolated to the `f09` grid, inline.

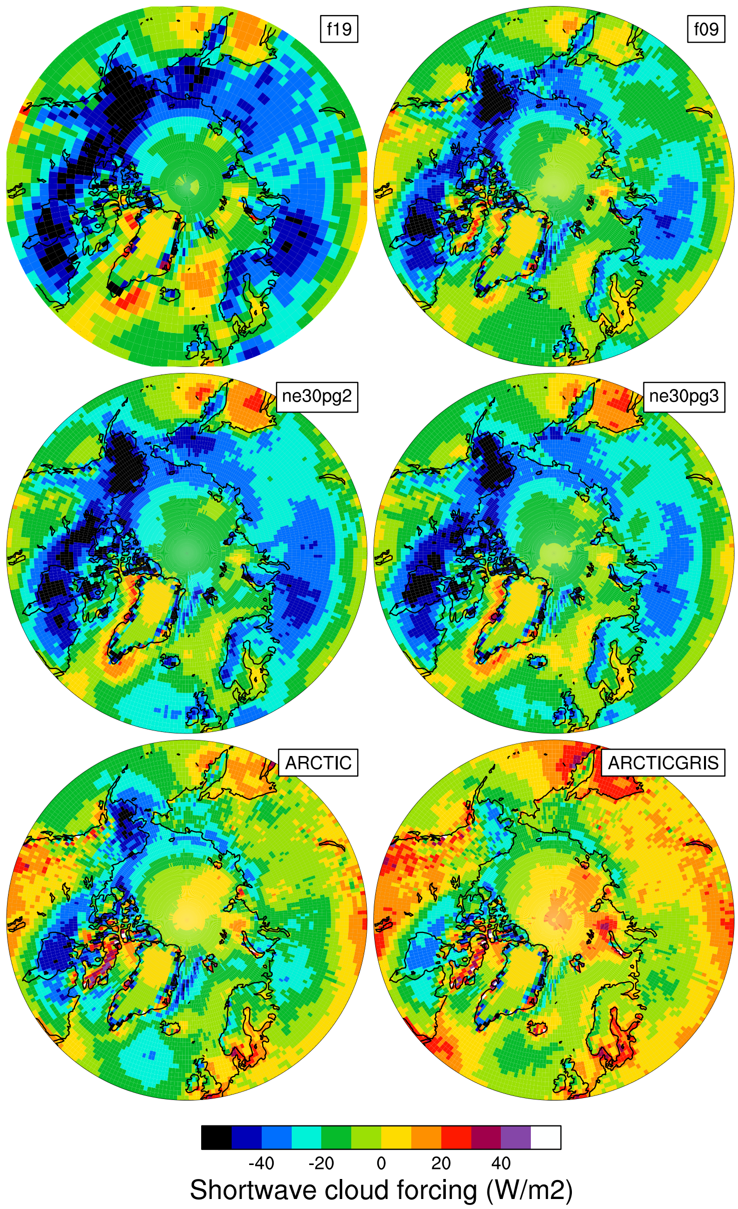

Figure 7. 1979-1998 Northern Hemisphere summer shortwave cloud forcing bias, relative to the CERES-EBAF gridded dataset. Shortwave cloud forcing is defined as the difference between all-sky and clear-sky net shortwave fluxes at the top of the atmosphere. Differences are computed after mapping all model output to the 1° CERES-EBAF grid.

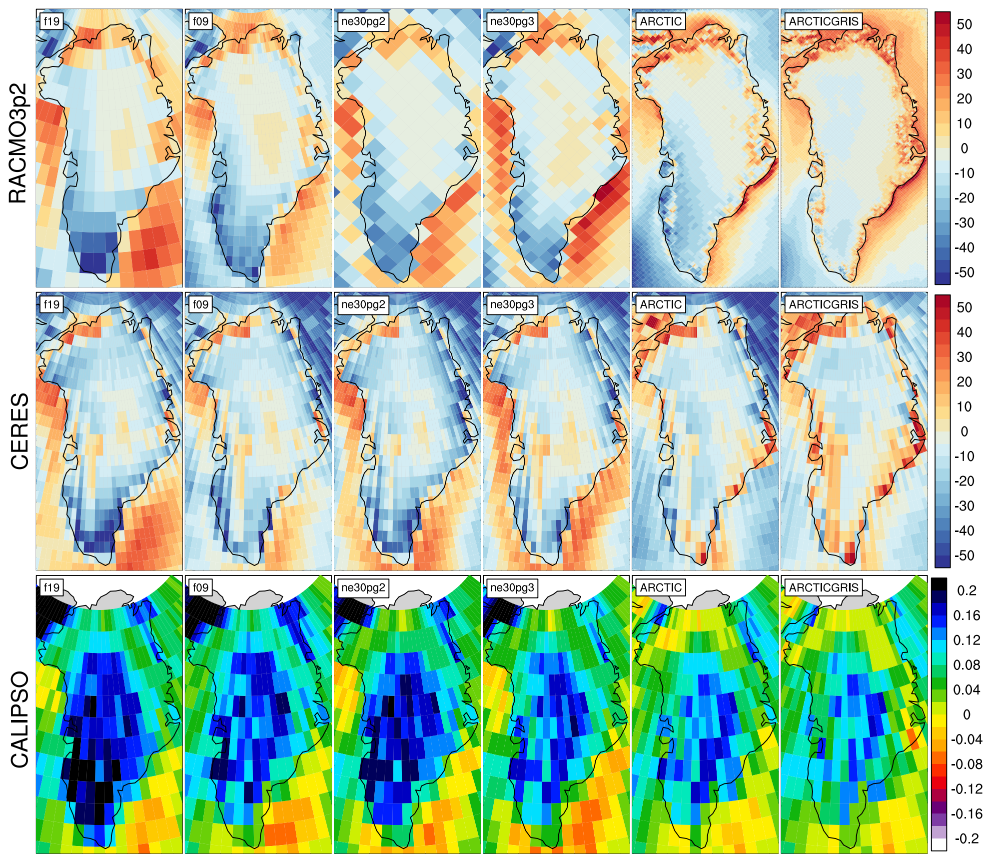

Figure 8. 1979-1998 northern hemisphere summer (top) total cloud fraction bias relative to the CALIPSO dataset and incident shortwave radiation bias (W/m2), computed as the difference (middle) from CERES, and (bottom) RACMO2.3p2 dataset. The CALIPSO and CERES differences are found by mapping the model output to the 1° grid, and differences in the bottom panel are computed after mapping the RACMO2.3p2 dataset to the individual model grids. Note that the averaging period for the CALIPSO-GOCCP and CERES-EBAF panels, 2006-2017 and 2003-2020, respectively, are different from the averaging period for the model results.

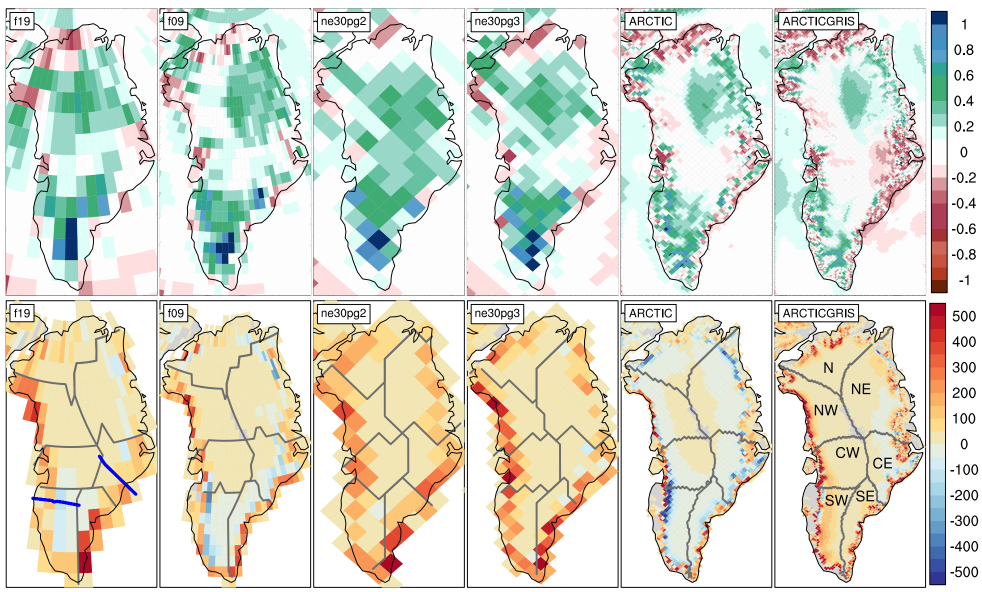

Figure 9. 1979-1998 (top) annual precipitation and (bottom) ice/snow melt biases relative to RACMO2.3p2, evaluated on the native model grids. The precipitation biases are expressed as fractional changes, whereas the melt biases are absolute changes (mm/yr). In the bottom panel, the Rignot and Mouginot (2012) basin boundaries are shown in grey for each model grid. Note that Figure 11 uses the basin boundaries for the two common ice masks, shown in the `f19` and `ne30pg2` panels, in computing the basin-scale integrals. Blue lines in the `f19` panel show the location of the two transects plotted in Figure 12..

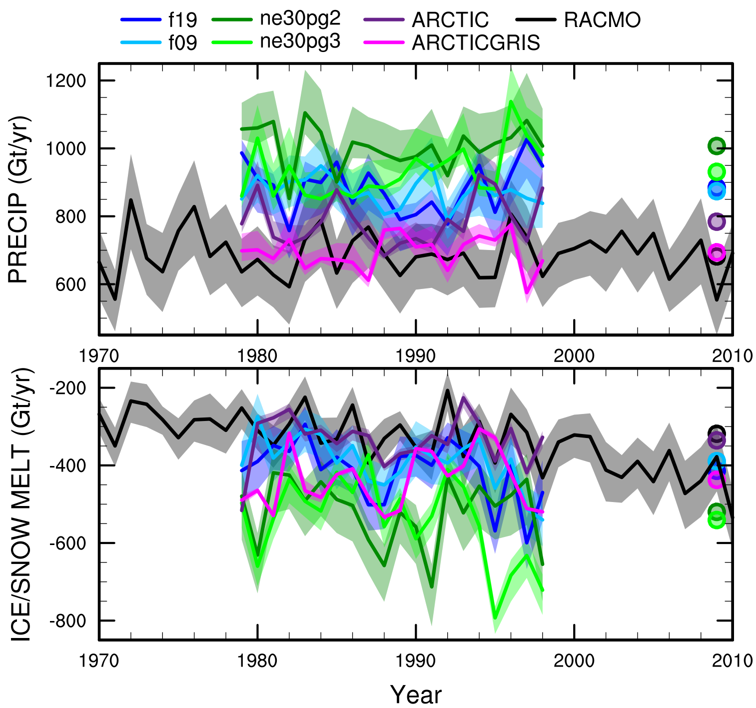

Figure 10. Time-series of annual (solid+liquid) precipitation (top) and annual runoff (bottom) integrated over the Greenland Ice Sheet for all six simulations and compared to the RACMO datasets. The time-series were generated using the common ice mask approach, which results in up to 4 ensembles, with the mean value given by the solid line and shading spanning the extent of the ensemble members.

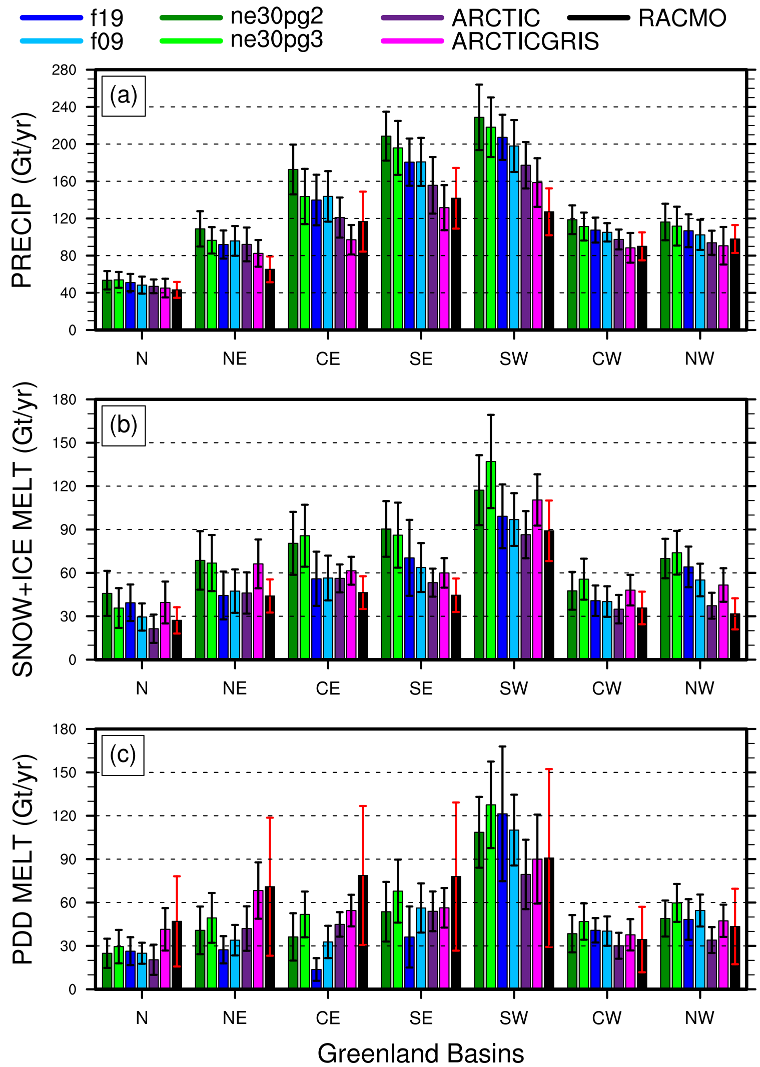

Figure 11. 1979–1998 basin integrated components of the SMB; (top) precipitation, (middle) ice/snow melt and (bottom) ice/snow melt estimated from the PDD method. Whiskers span the max/min of the four ensemble members generated from the common-ice-mask approach. Basin definitions are after Rignot and Mouginot (2012), and are found on the common ice masks using a nearest neighbor approach, and shown in Figure 9.

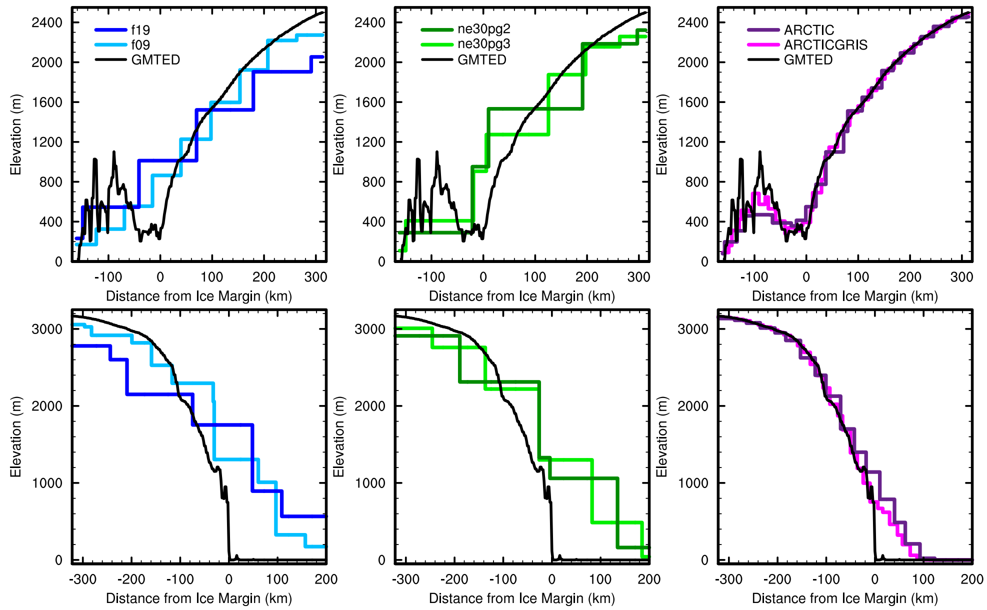

Figure 12. Model surface elevation along the (top) K-transect, and (bottom) a transect spanning the central dome down to the Kangerlussuaq glacier in southeast Greenland, for all model grids. The reference surface (GMTED) is a 1 km surface elevation dataset used for generating the CAM topographic boundary conditions.

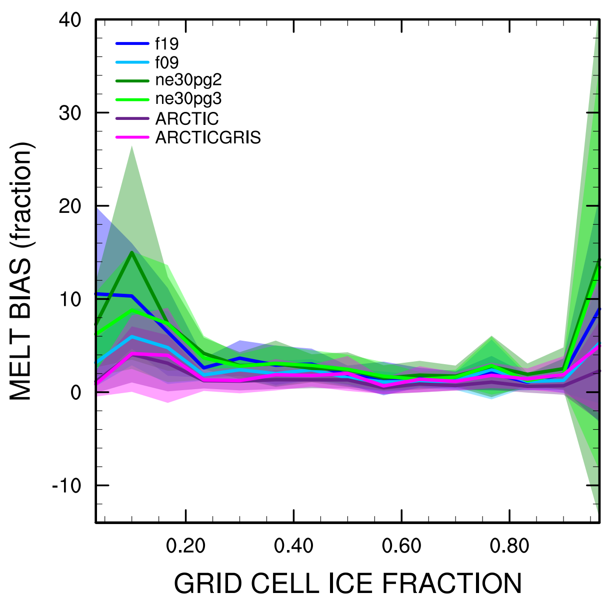

Figure 13. Fractional melt bias over the GrIS, computed relative to the RACMO datasets using the common ice mask approach, and conditionally sampled by grid cell ice fraction provided by the common ice masks. Solid lines are the mean of the distribution with ± one standard deviation expressed by shading.

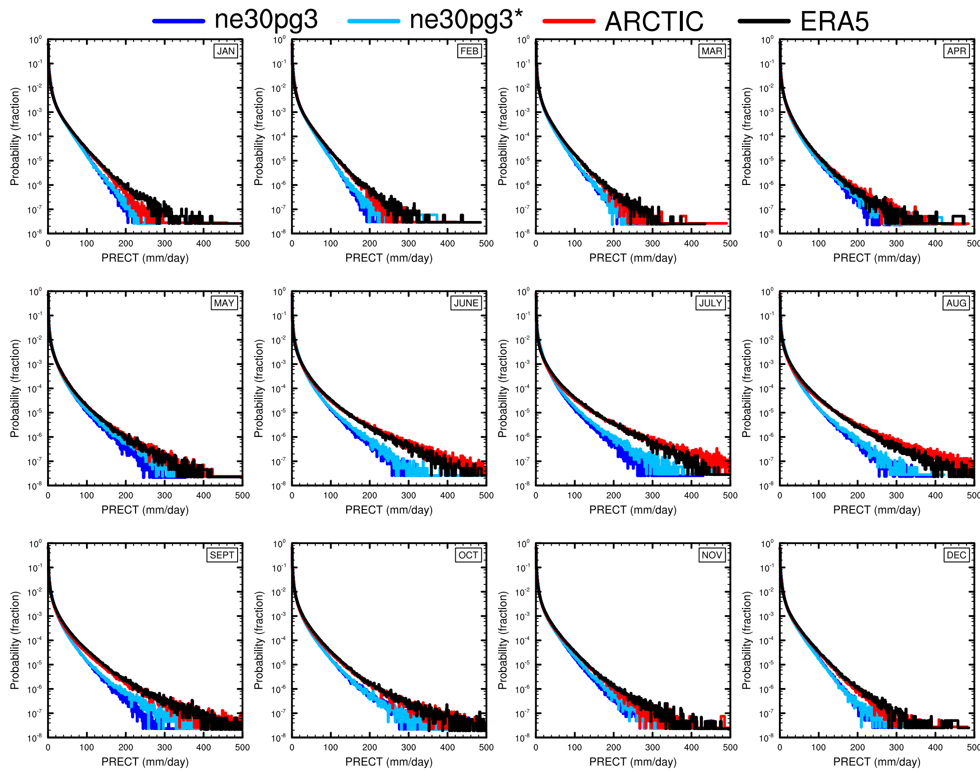

Figure 14. PDFs of the total precipitation rate associated with tracked storms, by month, in the `ne30pg3`, `ne30pg3`* and `Arctic` runs, and compared with the ERA5 dataset.

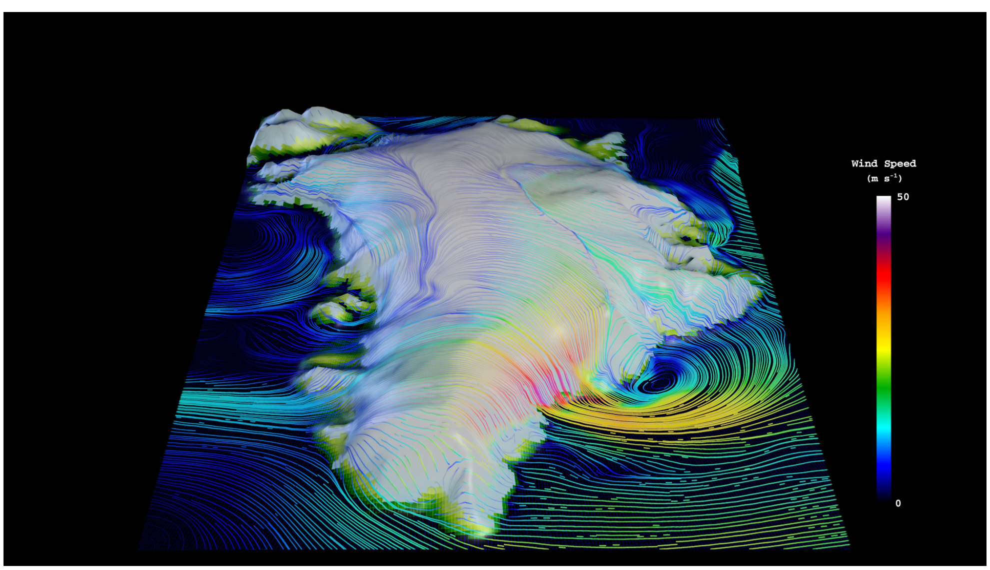

Figure 15. Snapshot of the lowest model level streamlines from the `Arctic` $-$ `GrIS` visualization, with color shading denoting the wind magnitude.

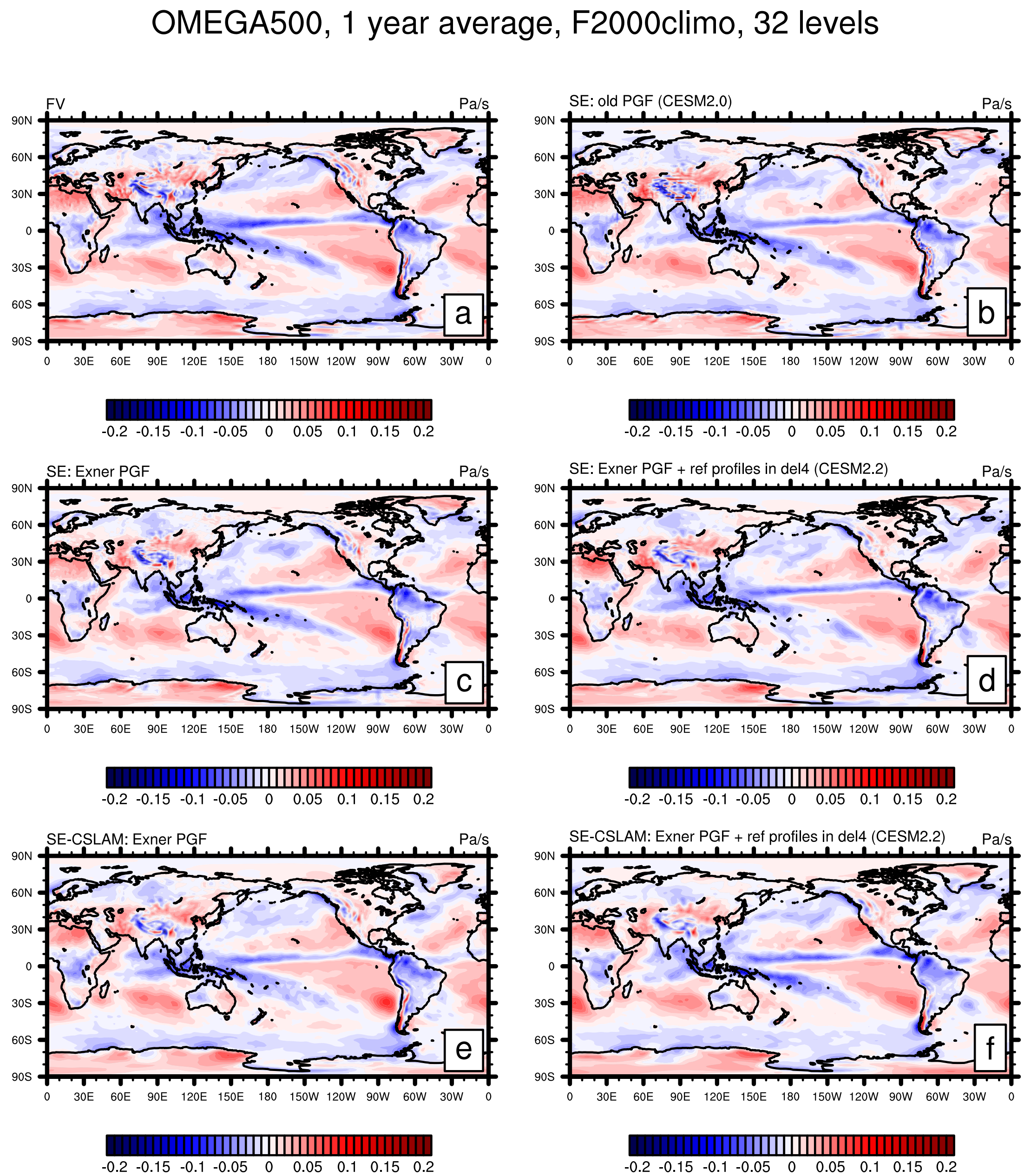

Figure A1. One year averages of vertical pressure velocity at 500hPa (`OMEGA500`) using (a) CAM-FV (Finite-Volume dynamical core) and (b-f) various versions of the spectral-element (SE) dynamical core at approximately 1° horizontal resolution and using 32 levels. (b) is equivalent to the CESM2.0 version of the SE dynamical core using the "traditional"/"old" discretization of the pressure-gradient force (PGF). Plot (c) is equivalent to configuration (b) but using the Exner form of the PGF. Plot (d) is the same as configuration (c) but also subtracting reference profiles from pressure and temperature before applying hyperviscosity operators (which is equivalent to the CESM2.2 version of SE in terms of the dynamical core). Plots (e) and (f) are equivalent to (c) and (d), respectively, by using the SE-CSLAM (`ne30pg3`) version of the SE dynamical core (i.e. separate quasi-uniform physics grid and CSLAM transport scheme).

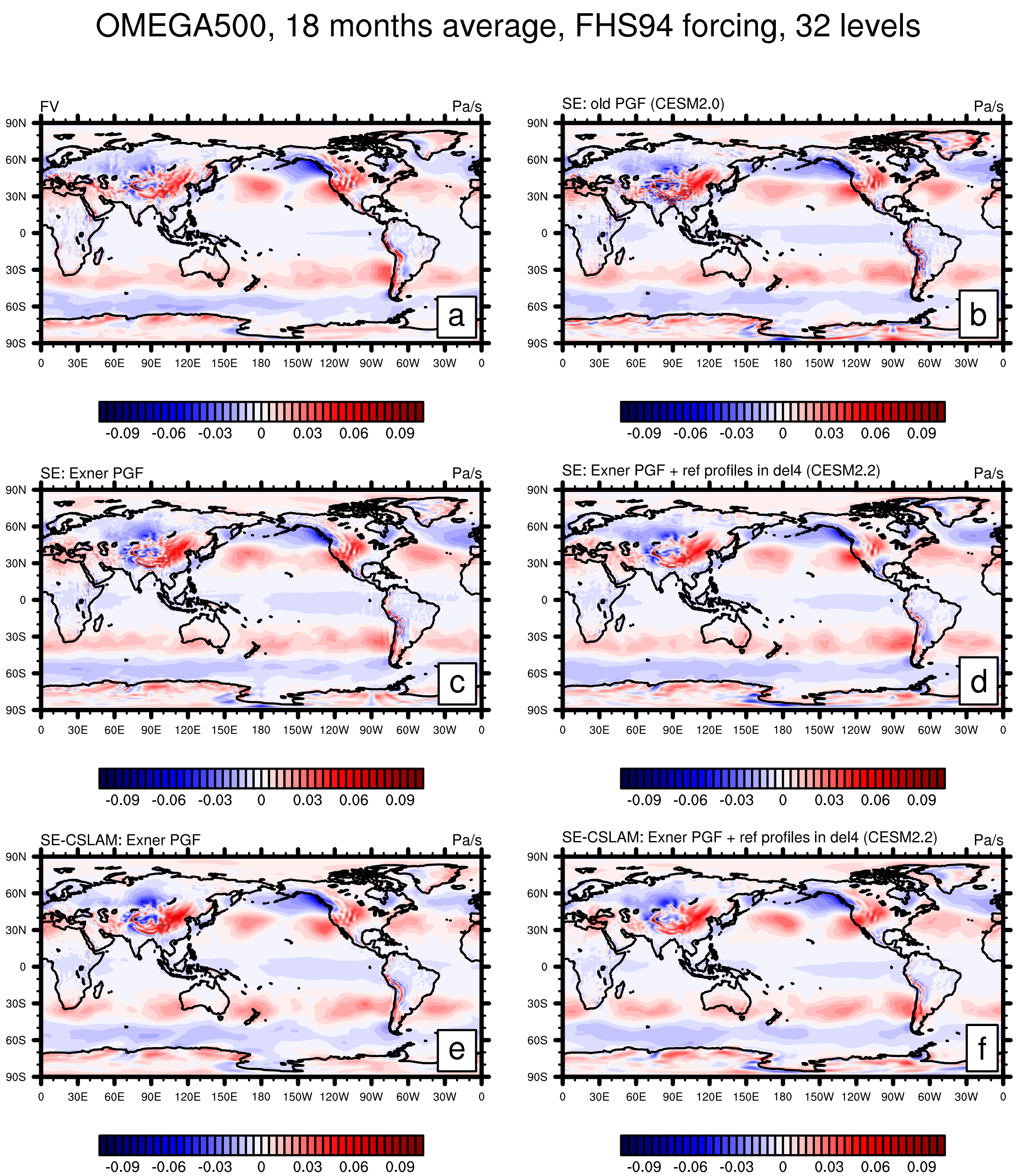

Figure A2. Same as Figure A1 but using modified Held-Suarez forcing and the average is over 18 months (excl. spin-up).