The Fully Coupled Regionally Refined Model of E3SM Version 2: Overview of the Atmosphere, Land, and River Results

Qi Tang[1], Jean-Christophe Golaz[1], Luke P. Van Roekel[2], Mark A. Taylor[3], Wuyin Lin[4], Benjamin R. Hillman[3], Paul A. Ullrich[5], Andrew M. Bradley[3], Oksana Guba[3], Jonathan D. Wolfe[2], Tian Zhou[6], Kai Zhang[6], Xue Zheng[1], Yunyan Zhang[1], Meng Zhang[1], Mingxuan Wu[6], Hailong Wang[6], Cheng Tao[1], Balwinder Singh[6], Alan M. Rhoades[7], Yi Qin[1], Hong-Yi Li[8], Yan Feng[9], Yuying Zhang[1], Chengzhu Zhang[1], Charles S. Zender[10], Shaocheng Xie[1], Erika L. Roesler[3], Andrew F. Roberts[2], Azamat Mametjanov[9], Mathew E. Maltrud[2], Noel D. Keen[7], Robert L. Jacob[9], Christiane Jablonowski[11], Owen K. Hughes[11], Ryan M. Forsyth[1], Alan V. Di Vittorio[7], Peter M. Caldwell[1], Gautam Bisht[6], Renata B. McCoy[1], L. Ruby Leung[6], and David C. Bader[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA
[2]Los Alamos National Laboratory, Los Alamos, NM, USA
[3]Sandia National Laboratories, Albuquerque, NM, USA
[4]Brookhaven National Laboratory, Upton, NY, USA
[5]Department of Land, Air and Water Resources, University of California, Davis, CA, USA
[6]Pacific Northwest National Laboratory, Richland, WA, USA
[7]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[8]Department of Civil and Environmental Engineering, University of Houston, TX, USA
[9]Argonne National Laboratory, Lemont, IL, USA
[10]Departments of Earth System Science and Computer Science, University of California, Irvine, CA, USA
[11]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Correspondence: Qi Tang (tang30@llnl.gov)

Abstract. This paper provides an overview of the United States (US) Department of Energy's (DOE's) Energy Exascale Earth System Model version 2 (E3SMv2) fully coupled Regionally Refined Model (RRM) and documents the overall atmosphere, land, and river results from the Coupled Model Intercomparison Project 6 (CMIP6) DECK (Diagnosis, Evaluation, and Characterization of Klima) and historical simulations – a first-of-its-kind set of climate production simulations using RRM. The North American (NA) RRM (NARRM) is developed as the high-resolution configuration of E3SMv2 with the primary goal of more explicitly addressing DOE's mission needs regarding impacts to the US energy sector facing Earth system changes. The NARRM features finer horizontal resolution grids centered over NA, consisting of 25→100 km atmosphere and land, 0.125° river routing model, and 14→60 km ocean and sea ice. By design, the computational cost of NARRM is ∼3x of the uniform low-resolution (LR) model at 100 km but only ∼10-20% of a globally uniform high-resolution model at 25 km.

A novel hybrid timestep strategy for the atmosphere is key for NARRM to achieve improved climate simulation fidelity within the high-resolution patch without sacrificing the overall global performance. The global climate, including climatology, time series, sensitivity, and feedback, is confirmed to be largely identical between NARRM and LR as quantified with typical climate metrics. Over the refined NA area, NARRM is generally superior to LR, including for precipitation and clouds over the contiguous US (CONUS), summertime marine stratocumulus clouds off the coast of California, liquid and ice phase clouds

near the North polar region, extratropical cyclones, and spatial variability in land hydrological processes. The improvements over land are related to the better resolved topography in NARRM, whereas those over ocean are attributable to the improved air-sea interactions with finer grids for both atmosphere and ocean/sea ice. Some features appear insensitive to the resolution change analyzed here, for instance the diurnal propagation of organized mesoscale convective systems over CONUS, and the warm-season land-atmosphere coupling at the Southern Great Plains. In summary, our study presents a realistically efficient approach to leverage the fully coupled RRM framework for a standard Earth system model release and high-resolution climate production simulations.

1 Introduction

Global Earth system models (ESMs) are fundamental tools for understanding the past evolution of the climate system and projecting future climate changes under various anthropogenic scenarios. High horizontal resolution simulations on climate scales have been recognized as one of the increasingly important directions of ESM development in recent years (Demory et al., 2014; Haarsma et al., 2016). Compared to low-resolution models, high-resolution models show superior fidelity in representing both the large-scale circulation (e.g., meridional ocean heat transport) (Griffies et al., 2015) and small-scale processes (e.g., clouds and streamflow) (Haarsma et al., 2016, and references therein). More importantly, simulations with enhanced horizontal resolution exhibit improved skills in capturing regional climate change signals and facilitating process-level studies, which provide a crucial basis for assessing the impacts of climate extremes with augmented societal implications. However, fine resolution and multi-century simulations (with ensembles) are competing requirements for climate experiments due to limited computational and human resources. This conflict will likely continue to challenge the climate modeling community as evidenced by the fact that more than three times (72 vs. 23) as many model sources (including different versions of the same model) have published simulations at 100-km than at 25-km nominal resolutions in the current Coupled Model Intercomparison Project 6 (CMIP6) archive (https://esgf-node.llnl.gov/search/cmip6/, access date: 18/08/2022). This suggests that despite the commonly recognized benefits, not many modeling centers can afford to pursue routine high-resolution climate simulations.

The Energy Exascale Earth System Model (E3SM) project (Leung et al., 2020) is supported by the U.S. Department of Energy (DOE) with a primary goal of improving actionable predictions of Earth system variability and change by leveraging advanced DOE computational resources. Scientifically, E3SM development is motivated by modeling requirements in three overarching fields (i.e., water cycle, biogeochemistry, and cryosphere) to address the most critical DOE mission-related questions, such as water availability, wildfires, heat waves, and sea-level rise, which all pose challenges to the energy sector with climate change. High-resolution simulations are clearly more desirable to achieve these E3SM objectives since these processes have high spatiotemporal variability. However, uniformly increasing the grid size for climate production simulations is not an easy task even with DOE's world class high performance computing power. For example, the 25-km simulation is at least 32 times (16x more grid cells, 2x smaller physics timestep, and 4x smaller dynamical core timestep) more expensive than the 100-km version with the E3SM version 1 (E3SMv1) model (Caldwell et al., 2019), making high-resolution models not only much more computationally expensive to run but also to tune for skillful simulations. With these demands and limitations, a

multiscale approach is an attractive avenue for global ESMs to deliver high-resolution production simulations over target areas at a more economical cost.

50   The multiresolution method (Ringler et al., 2008; Leung et al., 2013), also known as regionally refined model (RRM) or variable-resolution (VR) model, was proposed to alleviate the computational burden of global ESMs by refining a fraction of the globe with higher resolution while keeping (without coarsening) the remaining area at lower resolution. The RRM method is a general tool for all major ESM components, such as atmosphere, land, ocean, and sea ice. With a careful design of the RRM mesh, the high-resolution grids can better represent fine-scale processes over an area of interest at a typical cost of only

55   ~10–20% of a comparable globally uniform high-resolution configuration. Compared to regional or nested climate models, global RRMs by design minimize the impacts from the lack of a two-way dynamical feedback between the refined area and the outside domain.

    Recently, an increasing number of studies have successfully applied the RRM technique in global ESMs to tackle a wide range of climate research themes from climatological statistics of idealized aquaplanet (Zarzycki et al., 2014) and mean climate

60   state of more realistic simulations (Sakaguchi et al., 2015, 2016; Gettelman et al., 2018; Tang et al., 2019) to complex terrain climate (Wu et al., 2017; Rhoades et al., 2018c; Rahimi et al., 2019; Bambach et al., 2021) and climate extremes (Huang and Ullrich, 2017; Rhoades et al., 2020a, b; Zarzycki et al., 2021; Reed et al., 2022; Xu et al., 2022). Others leveraged RRM to study specific aspects of climate, such as tropical cyclones (Zarzycki and Jablonowski, 2014, 2015; Hazelton et al., 2018), marine stratocumulus (Bogenschutz et al., 2023), snowpack (Rhoades et al., 2016, 2017), surface energy flux (Burakowski

65   et al., 2019), Greenland surface mass balance (van Kampenhout et al., 2019), irrigation impacts on regional climate (Huang and Ullrich, 2016), and land use and land cover change influence on land-atmosphere coupling and precipitation (Devanand et al., 2020). Lately, the RRM resolution has been pushed to a new limit for watershed-scale hydrology analysis (Xu and Di Vittorio, 2021) and cloud-resolving scale climate simulation (Liu et al., 2022).

    The RRM high-resolution results are robust for most places except the Inter-Tropical Convergence Zone (Rauscher et al.,

70   2013; Zarzycki et al., 2014), covering almost all typical climate regimes such as the contiguous US (CONUS) (Gettelman et al., 2018; Tang et al., 2019), the western (Rhoades et al., 2016; Huang et al., 2016; Huang and Ullrich, 2017; Rhoades et al., 2018c) and eastern US (Liu et al., 2022), South America (Sakaguchi et al., 2015, 2016; Bambach et al., 2021), Asia (Sakaguchi et al., 2016), East Asia (Liang et al., 2021), eastern China (Xu et al., 2021), the Tibetan Plateau (Rahimi et al., 2019), the Maritime Continent (Harris and Lin, 2014), Atlantic basin (Zarzycki et al., 2015), Southeast Pacific (Bogenschutz

75   et al., 2023), Greenland (van Kampenhout et al., 2019), and Arctic (Veneziani et al., 2022). Furthermore, the RRM capability in representing the high-resolution climate seems general for different models, including Variable-Resolution Community Earth System Model (VR-CESM) (e.g., Gettelman et al., 2018), E3SMv1 atmospheric model (EAMv1) (Tang et al., 2019), Model for Prediction Across Scales-Atmosphere (MPAS-A) (Hagos et al., 2013; Sakaguchi et al., 2015, 2016; Liang et al., 2021), the Geophysical Fluid Dynamics Laboratory finite-volume dynamical core on the cubed-sphere grid (Harris and Lin, 2013, 2014),

80   and ICOsahedral Non-hydrostatic Earth System Model (ICON-ESM) (Jungclaus et al., 2022).

    All of the aforementioned studies utilize RRMs for Atmospheric Model Intercomparison Project (AMIP-type) (Gates et al., 1999) simulations. Although these studies provide valuable experience and important knowledge about RRMs, modeling cen-

ters still face the question of how to transform such AMIP-type RRM achievements from individual scientific studies emphasizing specific climate aspects to a standard global ESM release version aiming at a much broader and general scope. At a minimum, the criteria of reasonable global climate should be satisfied for the RRM to be widely adopted for global ESM releases. Most previous AMIP-type RRM studies focus on the regional results within the refined grids without paying much attention to the outside domain. While this might be acceptable for targeted studies, one cannot release a global model without reasonable global results since such a model is expected to address the challenge of a long (multi-century) spin-up and demonstrate top-of-atmosphere (TOA) radiative balance in pre-industrial fully-coupled simulations. In addition, some physics parameterizations (e.g., deep convection) suffer from poor scale-awareness and hence require retuning as the model resolution increases (e.g., Xie et al., 2018). This implies significant model calibration efforts that modeling centers have to seriously consider when planning on releasing the RRM besides the low-resolution model. Furthermore, the better RRM global climate is not warranted by retuning based on our EAMv1 RRM experience. In the present study, building upon the EAMv1 RRM (atmosphere and land of 25→100 km horizontal resolution with the 25-km mesh over the CONUS) (Tang et al., 2019) plus the E3SMv2 lower resolution configuration (Golaz et al., 2022), we extend the RRM configuration to ocean and sea ice (see grids in Fig. A1) as a fully coupled RRM with fine meshes centered over North America (NA). We propose an innovative RRM strategy (see details in Section 2.1) to meet the criteria above with a minimal retuning effort and for the first time to deliver production climate simulations using a fully coupled RRM.

This paper focuses on the atmosphere, land, and river components of the E3SMv2 North American RRM (NARRM), while a companion paper (Van Roekel et al., 2023) overviews the NARRM ocean and sea ice. This paper is organized as follows. Section 2 describes the NARRM model, our hybrid timestep strategy for the atmospheric component, and key tools and tests used to create its atmospheric configuration. Section 3 summarizes the simulations performed in the present study and reports on the computational cost of the NARRM historical simulation relative to its lower-resolution (LR) counterpart. Analyses of model results start at the global scale in Section 4 and then shift to the high-resolution NA region in Section 5 for atmosphere, land and river, and land-atmosphere interactions. Conclusions and discussions are presented in Section 6.

2 Model description

Except for the mesh and mesh-related settings, E3SMv2 LR and NARRM essentially have the same atmosphere, land, and river components. They are upgraded from E3SMv1 and briefly described here. In the E3SMv2 atmosphere model (EAMv2), the dynamical core uses the High Order Method Modeling Environment (HOMME) package (Dennis et al., 2005, 2011; Evans et al., 2013) on the spectral element grid (Taylor and Fournier, 2010). HOMME has been updated to use a potential temperature formulation of the equations with a more accurate pressure gradient (Taylor et al., 2020; Herrington et al., 2022) and a new interpolation semi-Lagrangian scheme (Islet) for passive tracer transport (Bradley et al., 2022). The physics operates on a separate finite volume grid (Hannah et al., 2021), which has 4/9 as many columns as the corresponding spectral element grid (see Table 1) and hence runs about 2x faster than it would on the spectral element grid. The physics parameterization updates include the Cloud Layers Unified By Binormals scheme (Golaz et al., 2002; Larson, 2017) for subgrid turbulent transport and

cloud macrophysics; the Zhang-McFarlane (ZM) deep convection scheme (Zhang and McFarlane, 1995) with a new trigger method (Xie et al., 2019); gravity wave parameterizations following Richter et al. (2010) with additional modifications (Beres et al., 2004; Richter et al., 2019); the O3v2 package (Tang et al., 2021) for the prognostic stratospheric ozone; and the four-mode version of Modal Aerosol Module (MAM4) (Liu et al., 2016; Wang et al., 2020) with an updated treatment of dust aerosol (Feng et al., 2022). The same set of EAM physics parameters is used in the LR and NARRM simulations analyzed here. The LR grid is a quasi-uniform $1°$ cubed sphere grid with an average grid spacing of $\sim$100 km. The NARRM grid has an average grid spacing of $\sim$25 km over North America, transitioning to match the $\sim$100 km cubed-sphere grid over the rest of the globe. All simulations, except the idealized baroclinic wave simulations described later, utilize E3SM's standard 72 vertical levels (L72).

The E3SMv2 land model (ELMv2) runs on the same grid as the atmospheric physics. ELMv2 upgrades the prescribed vegetation distribution for better consistencies between land use and changes in plant functional types across platforms, and adopts the new shortwave radiation model SNICAR-AD (Dang et al., 2019) for snow and ice. The land use harmonization version 2f data (LUH2; https://luh.umd.edu/data.shtml) (Hurtt et al., 2020) are converted into E3SMv2 plant functional types with an updated version of the land use translator (Di Vittorio et al., 2014). The trajectory of land cover change has also been improved through better tracking of previous land use change. The E3SMv2 river routing model (Model for Scale Adaptive River Transport, MOSARTv2) utilizes the regular lat-lon grid ($0.5°$ for LR and $0.125°$ for NARRM). MOSARTv2 uses the kinematic wave method to route the runoff from ELM into the ocean model via an eight-direction-based river network (Li et al., 2013). More details about the E3SMv2 model are documented by Golaz et al. (2022).

2.1 EAM hybrid timestep strategy for RRM production simulations

In previous RRM studies, including the EAMv1 CONUS RRM (Tang et al., 2019), the atmospheric physics timestep is often chosen to be shorter than that of the globally uniform low-resolution model to match the highest resolution grids in the RRM. However, such treatment faces the challenge of satisfying the criteria above for the purpose of global climate production simulations. Mainly because the ZM deep convection scheme and other cloud parameterizations used by EAM are by design not scale-aware (Xie et al., 2018), if the EAM in NARRM used a shorter physics timestep than LR while keeping other physics parameters unchanged, the NARRM results on the unrefined portion of the mesh (covering a larger area than the refined portion) would not match the quality of the LR results and thus undermine the NARRM global performance. Furthermore, even if NARRM used the retuned high-resolution physics parameters along with the shorter physics timestep, we would still have degraded global simulation quality over the LR model (see Fig. A2 for the EAMv1 results). With all these considerations, in the present study when employing RRM for climate production campaigns, we opt for a hybrid timestep strategy in EAM, which is a combination of LR physics timestep and the high-resolution dynamics timesteps (see Table 1). In this way, NARRM retains much of the LR global climate characteristics with possible improvements at the refined area benefiting from the high-resolution dynamics. Moreover, this approach simplifies the RRM development as it naturally avoids further tuning the RRM beyond what was done for LR. This choice also ensures that the physics behaves as closely as possible between the LR and RRM simulations, to facilitate direct comparisons of their climates. It is worthwhile noting that the hybrid timestep strategy

150 is a practical choice before the scale-aware cloud parameterization becomes available. With the coarsened physics timestep, NARRM cannot take full advantage of resolved physics (e.g., updrafts) at 25 km.

2.2 EAM running on unstructured meshes

In EAM, the underlying grid is always treated as fully unstructured. EAM can run on any grid that represents a tiling of the sphere with quadrilateral elements. For quasi-uniform grids, EAM relies on cubed-sphere grids since these grids are simple to
155 construct. RRM grids are constructed by external tools described below. Internally, the code treats all these grids identically, the only difference being the various resolution-dependent parameters. For the dynamical core, these parameters consist of the many timesteps in the model (given in Table 1) and the hyperviscosity coefficient. The dynamical core timesteps are chosen to ensure stability of the model. For RRM grids, these timesteps are chosen to match those that would be used in a global model with the same resolution as the highest resolution contained within the RRM. For the NARRM grid used here, which includes
160 refinement down to 25 km, we use the same timesteps as would be used by a global 25-km configuration of EAM.

Table 1. Column numbers and timesteps of the atmosphere component used in LR and NARRM simulations.

Model	Column #		Timesteps (s)				
	Dynamics	Physics	Dynamics				Physics
			Hyperviscosity	Dycore	Dycore Remap	Tracer	
LR	48,602	21,600	300	300	600	1,800	1,800
NARRM	130,088	57,816	75	75	150	450	1,800

For hyperviscosity, EAM relies on a resolution-aware tensor hyperviscosity formulation (Guba et al., 2014) applied on each model surface. The tensor coefficients vary spatially based on the two length scales of each spectral element (derived from the eigenvalues of the reference element map). This operator has a built in scaling of Δx^3 with strength controlled by a coefficient ν with units of s^{-1}. The tensor is designed to have the proper directional resolution dependence for highly distorted elements,
165 while matching the traditional constant-coefficient hyperviscosity on square elements. In EAMv2, we use the tensor hyperviscosity operator with $\nu = 3.4 \times 10^{-8} s^{-1}$ for all grids (cubed-sphere and RRM) and at all resolutions. The only exception is the LR $1°$ cubed-sphere grid, where for continuity with older simulations we continue to use the constant-coefficient hyperviscosity operator with $\mu = 1 \times 10^{15} m^4/s$. For a uniform degree p spectral element grid with square elements, the tensor operator with coefficient ν is identical to a constant coefficient hyperviscosity operator with coefficient $\mu = \left(\frac{p}{2}\right)^3 R \Delta x^3 \nu$, where Δx is
170 the element edge length divided by p and R is the radius of the sphere. In EAM, we always use $p = 3$.

2.3 Key tools for the RRM configuration

A number of tools have been developed to streamline the workflow for EAM/ELM simulations on RRM grids. These are described as follows, in the approximate order they are employed:

- **The Spherical Quadrilateral Grid Generator (SQuadGen):** Generation of the atmosphere/land mesh is performed using SQuadGen (Ullrich, 2022; Guba et al., 2014). This tool translates a monochrome PNG image, which denotes the desired level of grid refinement on an equirectangular projection, to a mesh of refined quadrilaterals based on a cubed-sphere. The use of quadrilaterals is by necessity for compatibility with the spectral element dynamcial core. Transition regions are managed using "paving", that is, using predefined patterns of quadrilaterals which enable transition between coarse resolution and fine resolution regions. Smoothing of the grid is performed via spring dynamics. The spectral elements of the NARRM grid produced with this procedure are shown in Fig. 1. In the spectral element method, each field is represented by polynomials up to degree 3 within each element. The resolution represented by each element (its average length divided by 3) is shown in Fig. 2.

- **TempestRemap:** The TempestRemap package (Ullrich and Taylor, 2015; Ullrich et al., 2016) is used to generate conservative, consistent, and monotone linear maps between fields stored as volume averages (i.e., updated using the finite volume methods) and fields stored as spectral elements (i.e., as coefficients of a set of basis functions). The generated maps require the construction of an "overlap mesh", which is the union of the source and target face, the generation of an approximate map, and subsequent projection of the approximate map onto the linear space of conservative, consistent, and (optionally) monotone maps.

- **Topography generation:** To generate topography and associated surface roughness fields on the NARRM grid, we rely on the tool chain described in Lauritzen et al. (2015) combined with a topography smoothing tool included with HOMME. The use of HOMME's topography smoothing tool ensures that the smoothing is done with the same discrete Laplace operator used internally in the dynamical core.

- **NetCDF Operators (NCO):** NCO consists of a number of command-line tools that enable manipulation of netCDF files (Zender, 2008). The tools include variable extraction, remapping, and spatial and temporal averaging. Provenance information is preserved within the netCDF files to enable scientific reproducability.

2.4 Idealized test

Before running long coupled NARRM simulations, we first evaluate the dynamical core settings for the NARRM grid using a baroclinic instability test case. This test case establishes that the dynamical core behaves as expected in an idealized setting: the timesteps are stable, the model can capture high resolution features in the high resolution region, and the presence of the high resolution and mesh transition regions do not negatively impact the large scale behavior. For this evaluation, we use an extension of the dry baroclinic wave test case by Ullrich et al. (2014) with two idealized, analytically-prescribed mountains (Hughes and Jablonowski, 2023). The latter now serve as the trigger for baroclinic instability. The addition of the two mountains generates a flow with more energy at smaller scales as compared to the original test case, especially downstream of the mountains, making this an attractive test case for studying the impacts of resolution.
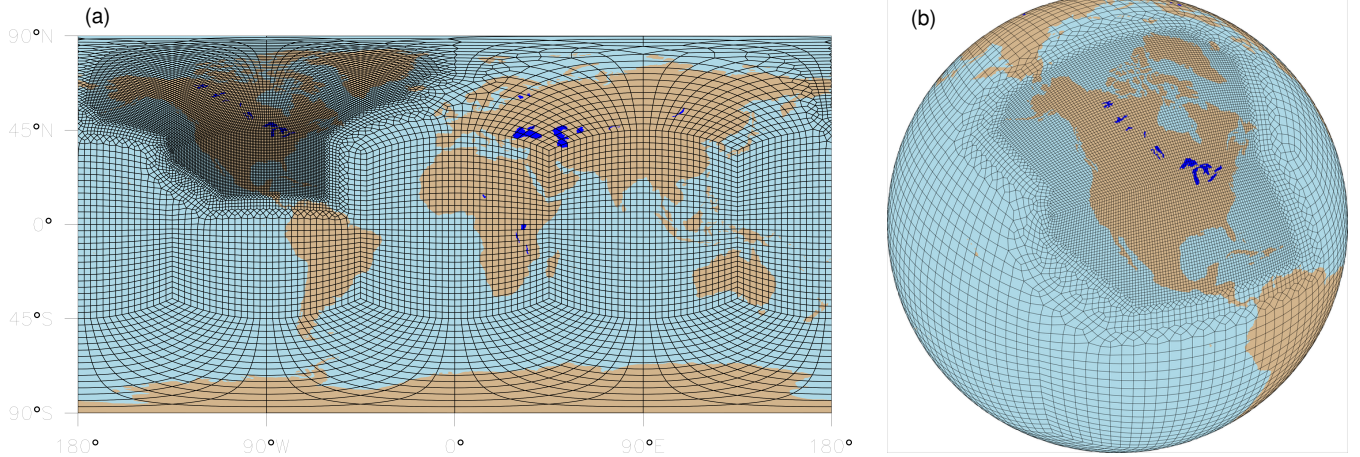
Figure 1. North American RRM (NARRM) grids for the atmosphere dynamical core shown in (a) a cylindrical equidistant projection and (b) an orthographic projection.

For this test case, we run simulations with three different horizontal grids, LR, NARRM, and high-resolution (HR), and 30 hybrid vertical levels (L30) which are specified in Appendix B of Reed and Jablonowski (2012). The LR and NARRM grids are as described above, and we add an HR grid. The HR grid is a global $0.25°$ grid which matches the high resolution region of the NARRM grid. All idealized runs use the same settings as in the full model (except L30 instead of L72), with HR and NARRM using identical timesteps since they both contain regions of $0.25°$ resolution. All simulations utilize the EAMv2 tensor hyperviscosity tuning with $\nu = 3.4 \times 10^{-8}$ and Δx^3 resolution scaling.

The test case is fully described in Hughes and Jablonowski (2023). We use the dry configuration and make one modification to the locations of the mountains. In particular, the center locations of the mountains are shifted longitudinally by $144°$ to the east in order to place the two mountains within the NARRM's high resolution region. The new center locations are therefore $144°$W and $76°$W. The peak height of the mountain ranges is 2000 m. Figure 2 illustrates the size and location of the mountains along with the NARRM mesh resolution while Fig. 3 shows the surface pressure at day 6 computed on the same mesh. The latter highlights the topographically generated baroclinic instability in the northern hemisphere.

Figure 4 shows contour lines of the 750 hPa temperature field after 6 days on all three grids. The plots are zoomed in over the region with the most activity shown in Fig. 3. We first compare the field in the NARRM's high resolution region with the HR result, and note the remarkable agreement between the two solutions (black contour lines) in the high resolution region (yellow color). The presence of high resolution in the NARRM simulation allows the model to capture features in that region with finer scales than can be captured by the LR simulation, as expected. Further downstream from the mountains at the right edge of the figure, the NARRM resolution has transitioned to match the LR resolution (blue color) and the scales captured by the NARRM solution are no longer as fine as they are in the HR solution. They are somewhat dissipated and fall between the LR and HR results. Thus the presence of the high resolution region in the NARRM grid improves some aspects of the solution in the low
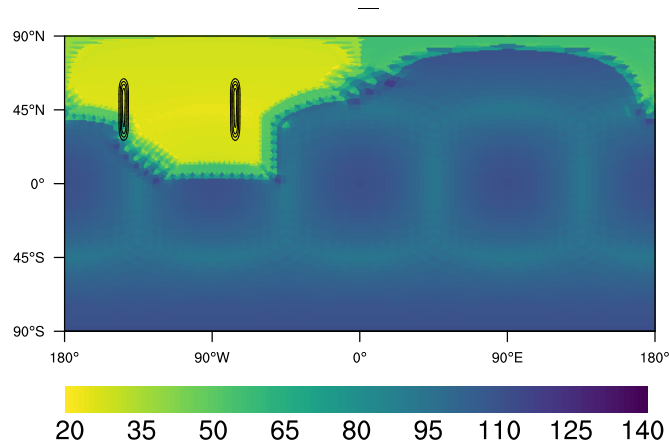
Figure 2. Contour lines of the topographic height with a peak amplitude of 2000 m overlaid on a map of the NARRM grid resolution (square root of element area). The resolution is ~25 km over North America (shown in yellow) transitioning to ~100 km over the rest of the globe (dark blue). The two mountains are mostly contained within the high-resolution region. In the low resolution region, the faint outline of an inscribed cube shows the slight nonuniformness of the 1° cubed-sphere grid used in that region.
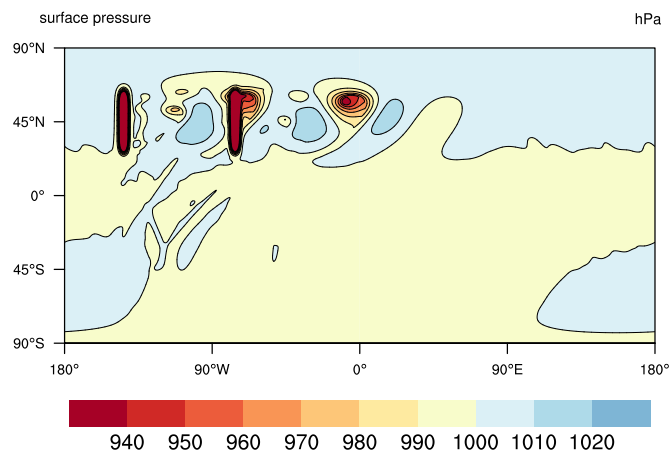
Figure 3. Contours of the surface pressure at day 6 showing the topographically triggered baroclinic instability in the northern hemisphere as computed on the NARRM grid. The instability has yet to be triggered in the southern hemisphere. The mountain height contours are overlaid. The colors saturate over the mountain ranges with minimum surface pressure values around 750–780 hPa (not shown).

225 resolution region. Finally, we note that there are no visible artifacts from the distorted elements in the mesh transition region. Examination of other fields, such as vorticity (not shown) demonstrate similar results.
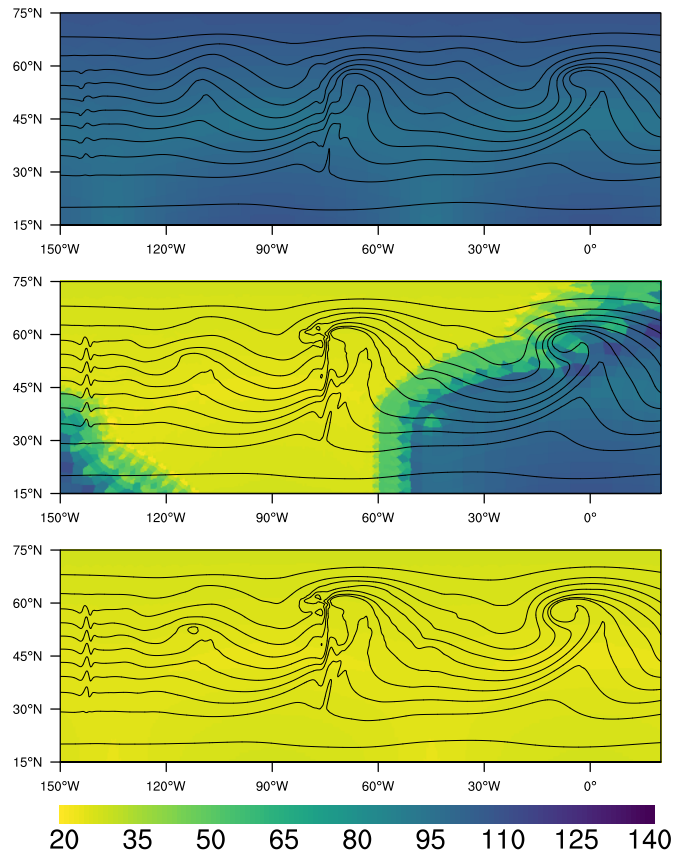
Figure 4. Contour lines of the 750 hPa temperature field on day 6 with contour intervals of 5° C. The temperature contours are overlaid on a map colorized by grid resolution. The data is plotted over a subset of the globe containing the mountains and most of the downstream region affected by the baroclinic instability. Results are shown from the LR grid (top), NARRM (middle), and HR grid (bottom). The NARRM grid shows the transition from high resolution (yellow, ~25 km) to low resolution (blue, ~100 km).

3 Simulations and computational cost

We perform a set of NARRM production simulations parallel to the LR version documented by Golaz et al. (2022) and following the same CMIP6 specifications. The LR and NARRM production simulations analyzed in the present study are
230 summarized in Table 2. These simulations consist of the CMIP6 Diagnosis, Evaluation, and Characterization of Klima (DECK) and historical simulations (Eyring et al., 2016), i.e., one pre-industrial control (*piControl*, 500 years), two idealized CO_2 runs (*1pctCO2* and *abrupt-4xCO2*, each 150 years), a five-member historical ensemble (*historical_N*, 1850–2014), and a

three-member Atmospheric Model Intercomparison Project (*amip*) type ensemble (*amip_N*, 1870–2014). Initial conditions are taken from January 1 of different years of *piControl* as indicated in Table 2 for *1pctCO2*, *abrupt-4xCO2*, and *historical_N*

235 simulations. *amip_N* simulations are initialized from the 1870 condition of corresponding *historical_N* simulations.

In order to estimate the effective radiative forcing of anthropogenic aerosols in LR and NARRM configurations, we perform pairs of nudged simulations with prescribed emissions of aerosols and their precursors for the present-day (PD, year 2010) and pre-industrial (PI, year 1850), which are taken from the CMIP6 emission data. Table 3 lists the nudged simulations used to assess the effective radiative forcing of anthropogenic aerosols. Horizontal winds in LR and NARRM are nudged towards

240 wind fields from their respective baseline simulations, with a relaxation time scale of 6h. These nudged simulations are 15-month long, with the first 3 months discarded as spin-up. Previous studies have shown that nudging the horizontal winds can help constrain the large-scale circulation in the model (Zhang et al., 2014; Sun et al., 2019; Tang et al., 2019), so that the anthropogenic aerosol effects can be determined with relatively short simulations (Zhang et al., 2022b, c).

Table 2. Summary of E3SMv2 LR (Golaz et al., 2022) and NARRM production simulations analyzed in this study. Numbers in parentheses indicate the simulation year numbers.

Label	Description	Period	Ens.	Initialization
Fully coupled				
(atmosphere, ocean, sea ice, land and river)				
piControl	Pre-industrial control	500 years	-	Pre-industrial spinup
1pctCO2	Prescribed 1% yr^{-1} CO_2 increase	150 years	1	*piControl* (101)
abrupt-4xCO2	Abrupt CO_2 quadrupling	150 years	1	*piControl* (101)
historical_N	Historical	1850-2014	5	*piControl* (101, 151, 201, 251, 301)
Prescribed SST and sea ice extent				
(atmosphere, thermodynamic sea ice, land and river)				
amip_N	Atmosphere with prescribed SSTs and sea ice concentration	1870-2014	3	*historical_N* (1870)

3.1 Computational performance

245 A sequence of performance benchmark simulations were run on the Argonne National Laboratory Chrysalis cluster. Chrysalis has 512 compute nodes. Each node has two AMD Epyc 7532 "Rome" 2.4 GHz processors. Each processor has 32 cores, for a total of 64 cores per node. Each node has 256GB 16-channel DDR4 3200MHz memory. The interconnect hardware is Mellanox HDR200 InfiniBand and uses the fat tree topology. The model code was compiled using Intel release 20200925 with GCC version 8.5.0 compatibility and run using OpenMPI 4.1.3 provided in the Mellanox HPC-X Software Toolkit.

Table 3. Nudged LR and NARRM atmospheric model simulations used in this study. All simulations are performed with prescribed sea surface temperature (SST) and sea ice concentration for year 2010. Nudging data are 6-hourly model output saved from the LR and NARRM free-running simulations (middle column). Due to the model instability problem with nudging application in RRM (with a relatively long time step), we use an alternative physics-dynamics coupling approach (see option "se_ftype = 1" in section 3.1 of Zhang et al. (2018)) for the NARRM nudged simulations. We find the impact of using different physics-dynamics coupling approaches on the global mean effective aerosol forcing estimate in LR to be small (difference < 0.05 W m^{-2}).

Label	Baseline simulation	Emission
Nudge_LR_PD	LR	2010
Nudge_LR_PI	LR	1850
Nudge_NARRM_PD	NARRM	2010
Nudge_NARRM_PI	NARRM	1850

250   The simulations are run with one MPI process per core and no OpenMP threading. Throughput values are computed using the maximum wall-clock time (minimum throughput) over all MPI processes; model initialization time is excluded. A throughput data point corresponds to one simulation run for 90 days. The input/output (I/O) configuration is identical to production simulations. At the end of 90 days, a restart file is written.
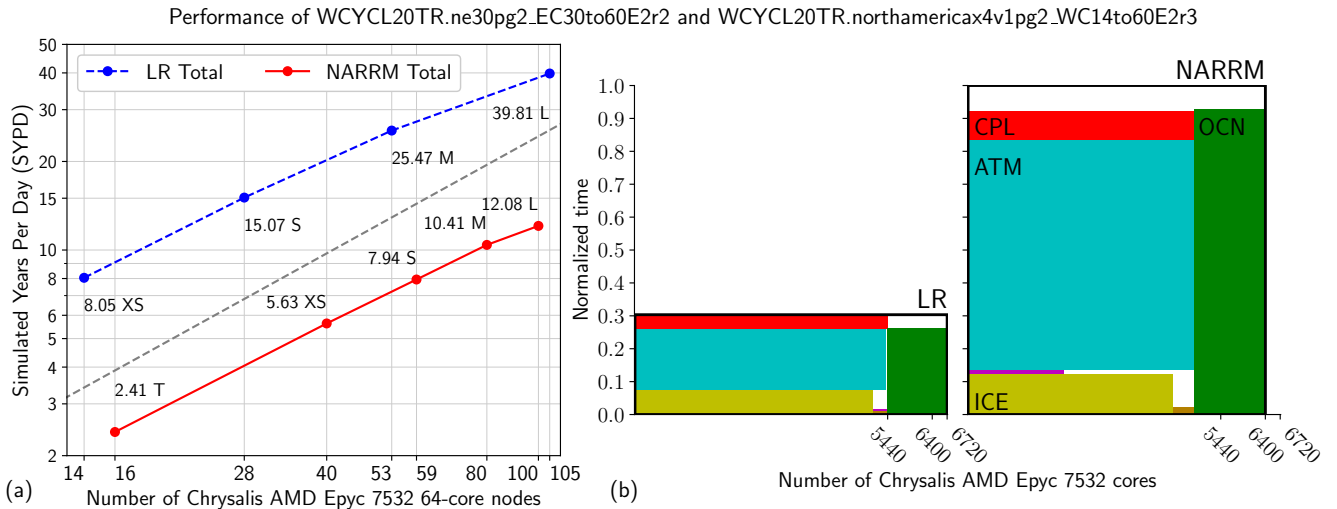
Figure 5. Performance of the LR and NARRM historical simulations. (a) Throughput vs. number of computer nodes. Each data point is annotated with its throughput in simulated years per day (SYPD) and computer resource configuration name. The dashed gray line shows the perfect-scaling slope. (b) Computational resource plots for the L process layouts. Each component has one rectangle. A rectangle has the area given by the product of normalized wall-clock time and number of cores, with the NARRM total time normalized to 1.0.

    Figure 5 summarizes the performance of the LR and NARRM *historical_N* simulations for several node counts and corre-
255 sponding process layouts with names T (NARRM only), XS, S, M, and L. Note that while the layout names are shared among
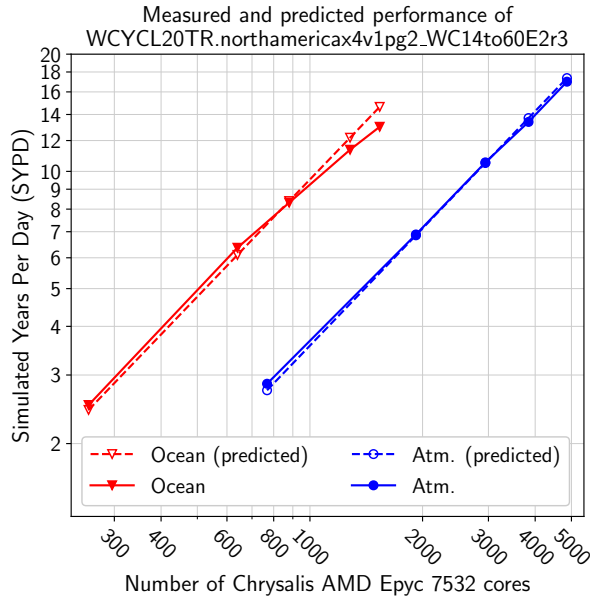
Figure 6. Performance of the atmosphere (Atm.) and ocean components of the NARRM historical simulation. Solid lines show measured performance. Dashed lines show the performance predicted by a simple model that uses the LR simulation with the XS process layout for input data; see the text for a description of the performance model.

models, the specific layout associated with a name differs among models. Each simulation's data point is annotated with its throughput in simulated years per day (SYPD) and process layout name. The highest throughput of the LR simulations is 39.81 SYPD. In Golaz et al. (2022, Fig. 2), the highest throughput is 41.89 for the same node count; *historical_N* simulations have additional forcings to compute relative to the *piControl* simulation used in Golaz et al. (2022, Fig. 2). The LR throughput falls

260 off from the perfect scaling slope faster than the NARRM throughput because the LR simulation has less work per node. For the L process layouts, and accounting for 105 vs. 100 nodes, the throughput factor difference is 3.14.

Figure 5b shows the wall-clock-time–resource product for each component for the L layouts. A rectangle's width is proportional to the number of cores the component uses; its height is proportional to the wall-clock time to simulate a fixed simulation period, with the time normalized so that the NARRM simulation has a total time of 1.0. The atmosphere (ATM), sea ice (ICE),

265 coupler (CPL), land (LND), and river runoff (ROF; LND and ROF are too small to label) components run on one set of nodes, while the ocean (OCN) component runs on another set. An unfilled rectangle having "LR" or "NARRM" at the top-right corner shows the total product. Because there is no global communication barrier between components run in sequence, the time value of each component is approximate, and, thus, the filled rectangles do not sum to the total time.

We can understand the NARRM component-level performance as a function of spatial and temporal discretization parameters

270 and one LR simulation to calibrate throughput. The LR calibration simulation should reflect that RRM simulations have a large amount of work per node; thus, we use the LR simulation run with the XS process layout, the left-most LR point in Fig. 5(a).

We focus on the two most expensive components, the atmosphere and ocean. We start with the ocean, whose performance is simpler to model. For simplicity, we write the formulas in terms of wall-clock time (w.c.t.) for a fixed simulation length, e.g., 90 days. The input measured datum is the top-level ocean-component (ocn) wall-clock time in the LR simulation run with the XS process layout, w.c.t.$_{\text{LR}}^{\text{ocn}}$. The input parameters are the number of computer cores (n_{core}) used in the LR (XS) and RRM (variable) simulations, the number of cells (n_{cell}) in each grid, and the time steps (Δt) in each simulation. For a fixed simulation length, the predicted ocean-component RRM performance is then

$$\text{w.c.t.}_{\text{RRM}}^{\text{ocn}} = \frac{(n_{\text{core}})_{\text{LR}}^{\text{ocn}}}{(n_{\text{core}})_{\text{RRM}}^{\text{ocn}}} \cdot \frac{(n_{\text{cell}})_{\text{RRM}}^{\text{ocn}}}{(n_{\text{cell}})_{\text{LR}}^{\text{ocn}}} \cdot \frac{(\Delta t)_{\text{LR}}^{\text{ocn}}}{(\Delta t)_{\text{RRM}}^{\text{ocn}}} \cdot \text{w.c.t.}_{\text{LR}}^{\text{ocn}}. \tag{1}$$

The performance model for the atmosphere is more complicated because it has two important time steps, one each for the dynamical core (dynamics) and the column parameterizations (physics). Thus, the factor accounting for model time steps is broken into two terms, one each for the physics and dynamics. The predicted atmosphere-component RRM performance is then

$$\text{w.c.t.}_{\text{RRM}}^{\text{atm}} = \frac{(n_{\text{core}})_{\text{LR}}^{\text{atm}}}{(n_{\text{core}})_{\text{RRM}}^{\text{atm}}} \cdot \frac{(n_{\text{cell}})_{\text{RRM}}^{\text{atm}}}{(n_{\text{cell}})_{\text{LR}}^{\text{atm}}} \cdot \left(\frac{(\Delta t_{\text{physics}})_{\text{LR}}^{\text{atm}}}{(\Delta t_{\text{physics}})_{\text{RRM}}^{\text{atm}}} \cdot \text{w.c.t.}_{\text{LR}}^{\text{atm physics}} + \frac{(\Delta t_{\text{dynamics}})_{\text{LR}}^{\text{atm}}}{(\Delta t_{\text{dynamics}})_{\text{RRM}}^{\text{atm}}} \cdot \text{w.c.t.}_{\text{LR}}^{\text{atm dynamics}} \right). \tag{2}$$

Figure 6 shows the results of these models, where wall-clock time and simulation length have been converted to throughput (SYPD). The solid lines show the measured throughput of each component as a function of number of computer cores. The dashed lines show the corresponding throughput values predicted by Eqs. 1 and 2. The single LR XS-layout ocean throughput value is used as the reference for the ocean and, similarly, the single LR XS-layout atmosphere throughput value is used as the reference for the atmosphere; these are the only measured data inputs to the performance models. The primary error in the performance model is not accounting for a fall-off in scaling at large core counts. Because this fall-off is small for the atmosphere and ocean components, these simple performance models are accurate and can be used to predict the cost of other model configurations. For example, a uniform high-resolution atmosphere model would use the ne120pg2 grid, which has $6 \cdot 120^2$ elements. Using the same time steps and number of vertical levels as in the NARRM configuration, which has 14454 elements, for fixed computational resources, the high-resolution atmosphere configuration's throughput would be $6 \cdot 120^2 / 14454 = 5.98$ times smaller than the NARRM configuration's throughput, where this factor is the quotient of the numbers of elements in each of the two grids.

4 Global climate

As described above, the RRM model is expected to simulate a global climate similar to the LR model for production simulation campaigns since most areas are still covered by the same LR grids. In this section, we will examine whether this is the case for the global mean climate, climate sensitivity, as well as climate feedback.

For the global climatology, we focus on the last three decades (years 1985–2014) of historical simulations when more observational datasets are available. Figure 7 provides an overall comparison of the global mean climate among LR (blue triangles), NARRM (red triangles), and CMIP6 (boxes and whiskers) models as quantified by the uncentered spatial root-mean-square
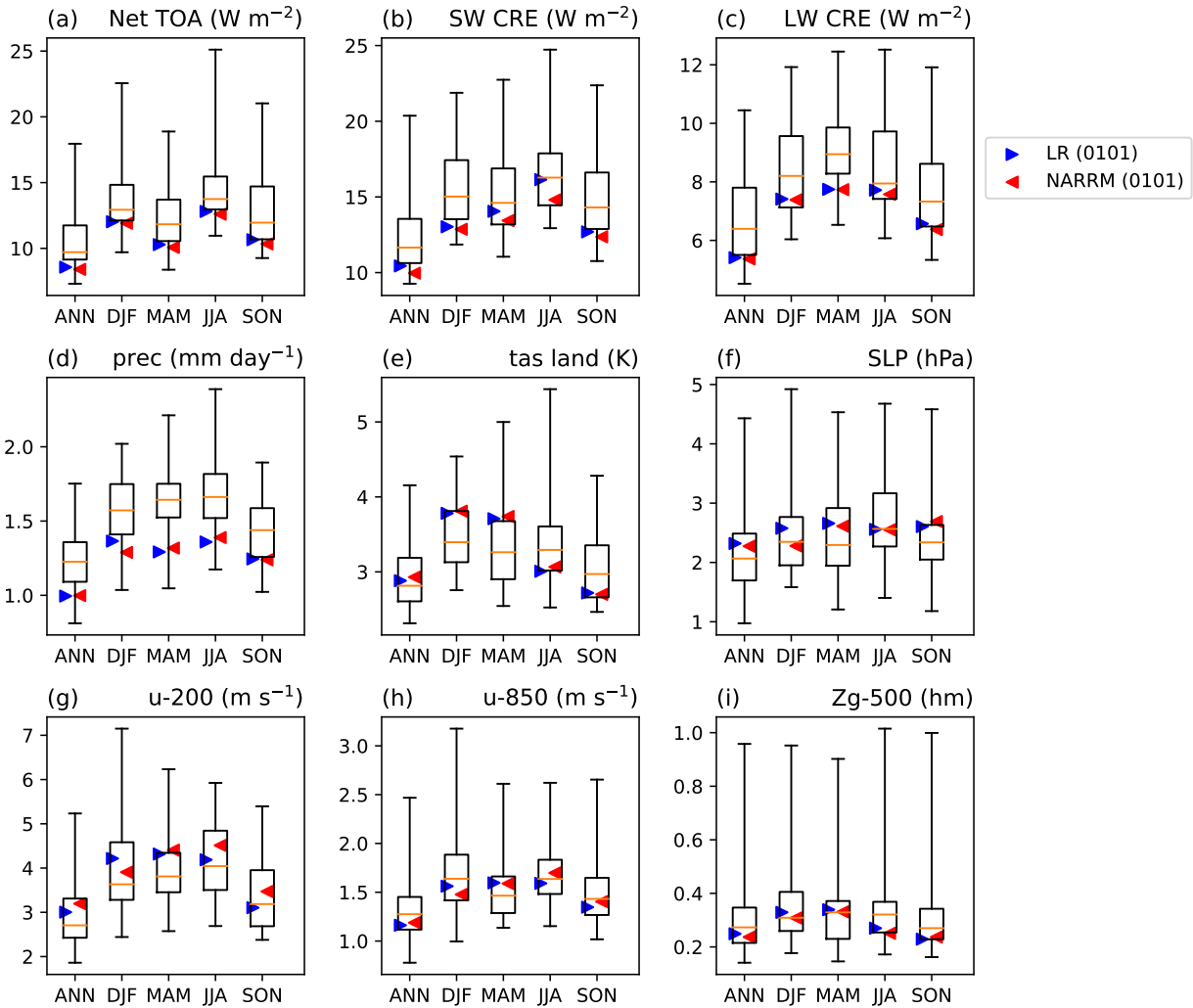
Figure 7. Comparison of the global spatial RMSE of model climatology (annual and seasonal averages of years 1985–2014) vs. observations with the E3SM Diags package (Zhang et al., 2022a). The model results are from the first historical member of E3SMv2 (0101) LR (blue triangles) and NARRM (red triangles), and 52 CMIP6 models (r1i1p1f1). The boxes and whiskers show 25th, 75th percentile, minimum and maximum RMSE of the CMIP6 ensemble. Quantities include (a) TOA net radiation flux, (b, c) TOA SW and LW cloud radiative effects, (d) precipitation, (e) surface air temperature over land, (f) sea-level pressure, (g, h) 200- and 850-hPa zonal wind, and (i) 500-hPa geopotential height. TOA = top-of-atmosphere; SW = shortwave; CRE = cloud radiative effects; LW = longwave; ANN = annual; DJF = December–February; MAM = March–April; JJA = June–August; SON = September–November; RMSE = root-mean-square error. The climatology of the observations and reanalysis data are calculated from CERES-EBAF Ed4.1 (Loeb et al., 2018) (2001–2014) for (a, b, and c), GPCP2.3 (Adler et al., 2018) (1985–2014) for (d) and ERA5 (Hersbach et al., 2020) (1985–2014) for (e, f, g, and h).
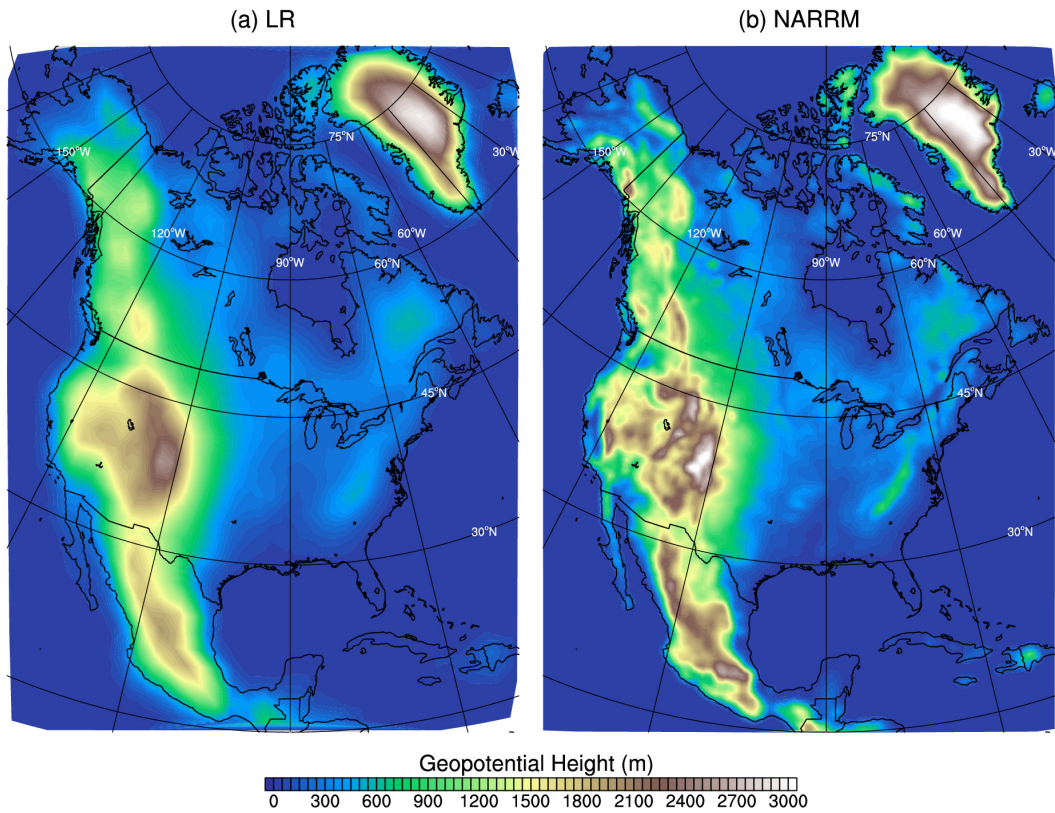
Figure 8. Surface geopotential height of (a) LR and (b) NARRM over North America.
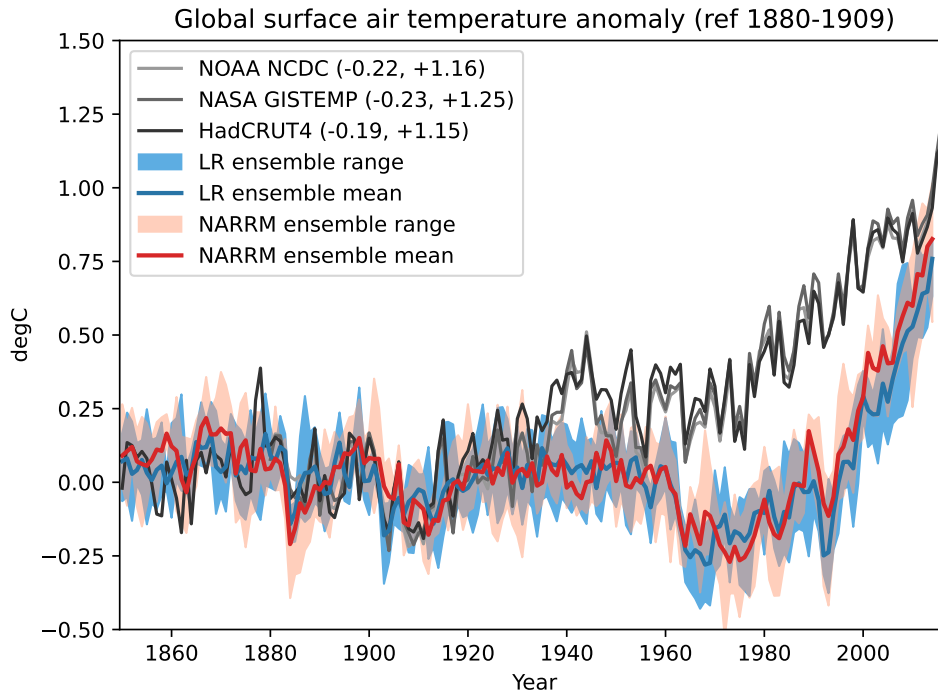
Figure 9. Time series of global annual mean surface air temperature anomalies from the ensemble mean of LR (blue) and NARRM (red) historical runs and observational datasets (grays) (National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC), National Aeronautics and Space Administration (NASA) GISTEMP, and HadCRUT4). The model ensemble minimum-maximum ranges are shaded, while the observational minimum and maximum numbers are labeled in the parentheses of legend.

error (RMSE) relative to the observations or reanalysis data. The RMSE numbers are calculated with the E3SM Diags package (Zhang et al., 2022a) for the first historical member (0101 for LR and NARRM, and r1i1p1f1 for CMIP6 models). Figure 7 clearly shows that NARRM and LR simulate very similar annual and seasonal averages. NARRM outperforms LR in the June-July-August (JJA) shortwave (SW) cloud radiative effect (CRE) partly because it better represents low clouds in NA (see Fig. 13 for the example at California). NARRM also simulates slightly better December-January-February (DJF) precipitation compared to LR, partly due to its improved topography (Fig. 8) and orographic precipitation in NA (see Fig. 12bdf). For other times (e.g., annual mean (Fig. A3)) NARRM and LR precipitation results are very similar. On the other hand, NARRM does not perform as well as LR for some other fields, such as the 200-hPa zonal wind in JJA and September-October-November (SON), which are associated with the increased positive biases in the tropical western Pacific and Amazon (not shown).

Figure 9 compares the long time series (years 1850–2014) of global annual average anomalies in the surface air temperature from the ensemble means of LR and NARRM historical simulations, and observational datasets (National Oceanic and Atmospheric Administration National Climatic Data Center (Smith et al., 2008; Zhang et al., 2015), National Aeronautics and
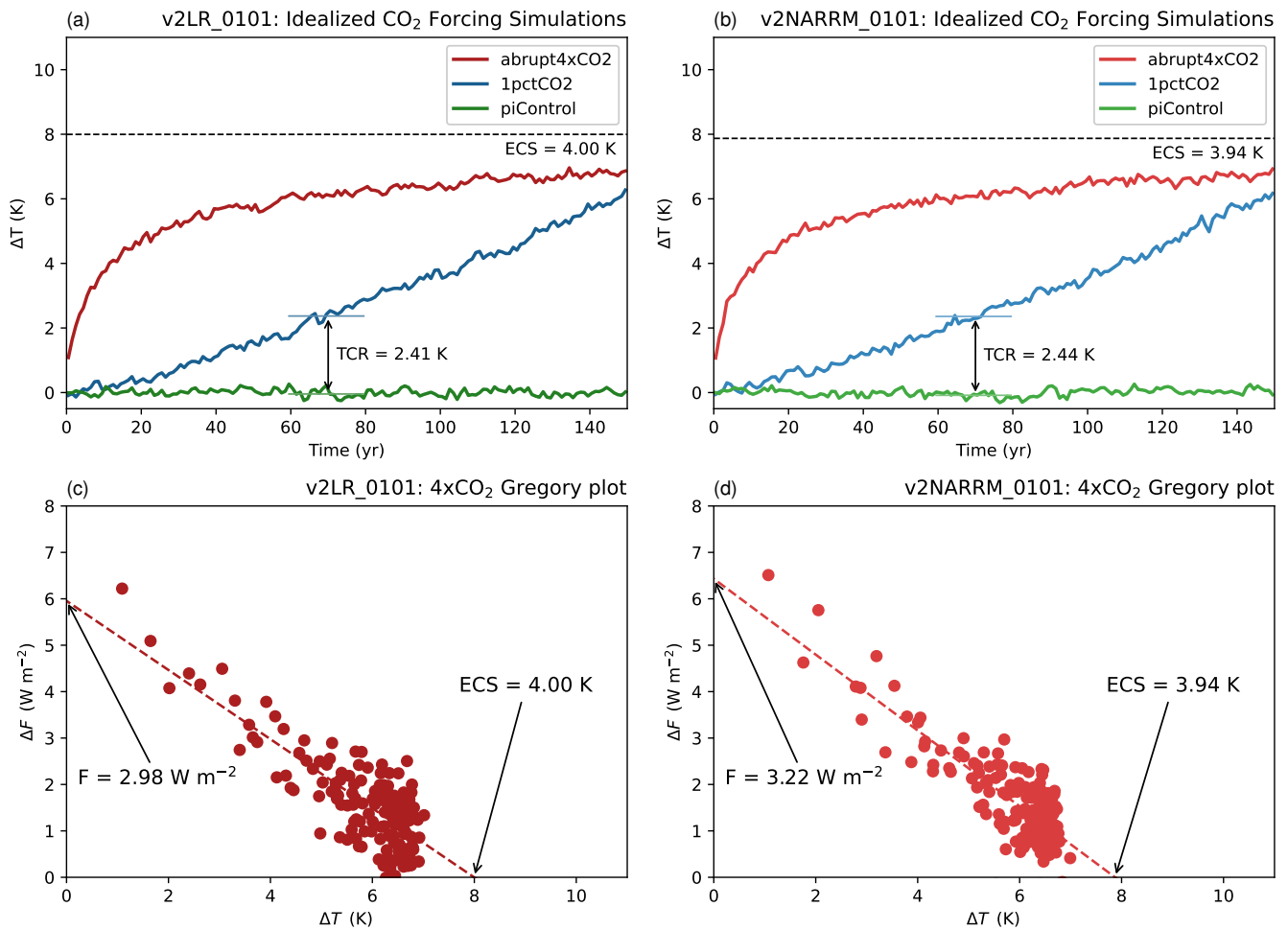
Figure 10. Comparison of climate sensitivities between LR (left) and NARRM (right) derived from idealized CO_2 forcing simulations. (a, b) time series of global annual mean surface air temperature anomaly from the following simulations, *abrupt-4xCO2* (red), *1pctCO2* (blue), and the control (*piControl*; green). The transient climate response (TCR) is computed as a 20-year average around time of CO_2 doubling (year 70). (c, d) Gregory regression plots. The estimated effective climate sensitivity (ECS) and effective 2x CO_2 radiative forcing (F) are as labeled.

315 Space Administration GISTEMP (GISTEMP Team, 2018; Hansen et al., 2010), and HadCRUT4 (Morice et al., 2012)). Over the whole period, NARRM tracks LR closely, including good agreement with observations until the 1930s and low biases afterwards, which are mainly attributed to too strong aerosol-related forcing and feedback (see Golaz et al., 2022, for details). This is further confirmed by the fact that the global mean effective radiative forcing of anthropogenic aerosols in NARRM and LR are very similar (-1.415 W/m^2 versus -1.421 W/m^2) as quantified by a pair of nudged simulations (see Section 5.1.2).

320 Following the CMIP6 DECK protocol (Eyring et al., 2016), we quantify the climate sensitivity and feedback with the abrupt quadrupling of CO_2 (*abrupt-4xCO2*) and the transient climate response (TCR) with a simulation forced by a 1% yr^{-1} CO_2 increase (*1pctCO2*) relative to the pre-industrial control simulation (*piControl*). The equilibrium climate sensitivity (ECS) is estimated with the linear regression of TOA radiation change against surface temperature change in a 150-year *abrupt-4xCO2* simulation (Gregory et al., 2004). The 2xCO_2 effective radiative forcing (ERF) is computed as the Y-intercept of the Gregory

325 plot divided by two, which measures the energy imbalance caused by doubling the atmospheric CO_2 concentration while keeping the surface temperature unchanged. TCR, which measures the response on shorter time scales, is derived based on its definition — the average surface temperature change in the 20-year period when the CO_2 concentration doubles from a *1pctCO2* experiment.

Figure 10 depicts the annual mean surface temperature change as a function of time and the Gregory plots from the idealized

330 CO_2 experiments with LR and NARRM. The differences in climate sensitivity between LR and NARRM are very subtle as quantified by both ECS (4.00 K vs. 3.94 K) and TCR (2.41 K vs. 2.44 K).

The regression slope in the Gregory plot (Fig. 10cd) denotes the total radiative feedback caused by the quadrupled CO_2 concentration. We further apply the radiative kernel method (Soden et al., 2008; Held and Shell, 2012) to decompose the total radiative feedback into non-cloud and cloud feedbacks. The cloud feedback is estimated by adjusting the cloud radiative

335 effect anomalies for non-cloud influences. Overall, NARRM shows a slightly larger ERF (3.22 W/m^2 vs. 2.98 W/m^2), which accompanied with the similar ECS produces a stronger negative total climate feedback in NARRM. The total climate feedback is -0.74 W/m^2/K for LR and -0.82 W/m^2/K for NARRM, which mainly relates to the slightly weaker positive SW cloud feedback in NARRM than in LR (see Fig. 22).

In summary, the results in this section confirm that NARRM with the hybrid timestep methodology simulates largely identical

340 global climate as its corresponding LR configuration and hence satisfies the necessary requirement (i.e., good global climate) of global RRM production simulations we proposed in the introduction.

5 North American results

In this section, we will zoom in over the refined region over North America (NA) and emphasize climate aspects most relevant to the E3SM water cycle scientific goals (Leung et al., 2020) as well as some weaknesses in LR revealed by Golaz et al. (2022).

345 The results will be described for the atmosphere, land, and river models, respectively. Moreover, we will analyze interactions between different components (i.e., land-atmosphere coupling), because these interactions are also expected to change with the resolution increase.

5.1 Atmosphere

5.1.1 Hydrology over the CONUS

350 First, we look at the overall atmospheric results over the CONUS (20°– 50°N, 65°– 125°W) by comparing the spatial RMSEs of the historical ensemble means between LR (blue triangles) and NARRM (red triangles) in Fig. 11. The same metric is used in Fig. 7 for the global results, but we adjust the variables to be more relevant to the CONUS. NARRM generally produces better (as quantified by smaller RMSE numbers) results than LR for these annual and seasonal climatologies, such as SW CRE (Fig. 11a), precipitation (Fig. 11c), and 200-hPa zonal wind (Fig. 11f). Because we have not retuned the physics of NARRM,

355 some deteriorations are expected, for example longwave (LW) CRE in DJF and March-April-May (MAM) (Fig. 11b).

Precipitation and clouds, which are obviously important quantities for the water cycle, are largely improved in NARRM compared to LR. The precipitation patterns are better captured by NARRM as observed at the Sierra Madre Occidental in JJA (Fig. 12 left column) and in the western US in DJF (Fig. 12 right column) due to the better resolved topography in NARRM (see Fig. 8). The poor representation of marine stratocumulus clouds is a long-standing problem that plagues many ESMs (e.g.

360 Bogenschutz et al., 2023). The underestimation of summertime low clouds (manifested as the excessive TOA shortwave CRE) in the California stratocumulus region are substantially improved with NARRM (Fig. 13). The improvement in this bias is likely due to a reduced bias in the simulated sea surface temperature (SST) in the coupled RRM. To constrain the impact of the RRM on the SST bias, we conduct two additional experiments: one where the LR ocean is coupled to the RRM atmosphere, and another where the RRM ocean is coupled to the LR atmosphere. The simulated SST bias averaged over years 51–100 of the

365 piControl is shown in Fig. 14. Comparison of Fig. 14 (a) and (b) shows a clear reduction in bias for the NARRM simulation. Figure 14 (c) and (d) illustrate that the regional refinement in the atmosphere (c) is primarily responsible for the reduction in SST bias, however, comparing to the NARRM result, we see that the bias is further reduced when regional refinement is included in both components, highlighting the advantage of coupled RRM over a single-component RRM. Further details are described in Van Roekel et al. (2023).

370 Another well-known issue of global ESMs is the poorly captured diurnal propagation of organized mesoscale convective systems (MCSs). Over CONUS, such MCSs originate from the front range of the Rockies in the afternoon and propagate eastward, manifesting as a nocturnal precipitation peak in the central US and contributing as much as half of the summertime rainfall in that region (Riley et al., 1987; Jiang et al., 2006). Both LR and NARRM simulate the summertime nocturnal rain peak in the central US (Fig. 15) because of the new convective trigger method for deep convection (Xie et al., 2019). However,

375 the magnitude is weaker and the area of nocturnal peak extends much larger (almost the whole east half of the US) than the observation, which could be caused by remaining propagation or convective trigger deficiencies. Nevertheless, NARRM reduces the underestimation of maximum diurnal cycle magnitude with an 80% greater value (5.61 mm/day vs. 3.10 mm/day) than LR on the same 100-km grids. The NARRM maximum can be as high as 6.71 mm/day on the 25-km grids, but still biases low compared to the observation (10.89 mm/day). This result suggests that a resolution of ∼25 km is not adequate to capture

380 the physics driving propagating MCSs, which probably require convection-permitting atmospheric simulations to achieve a good agreement with observations (Caldwell et al., 2021).
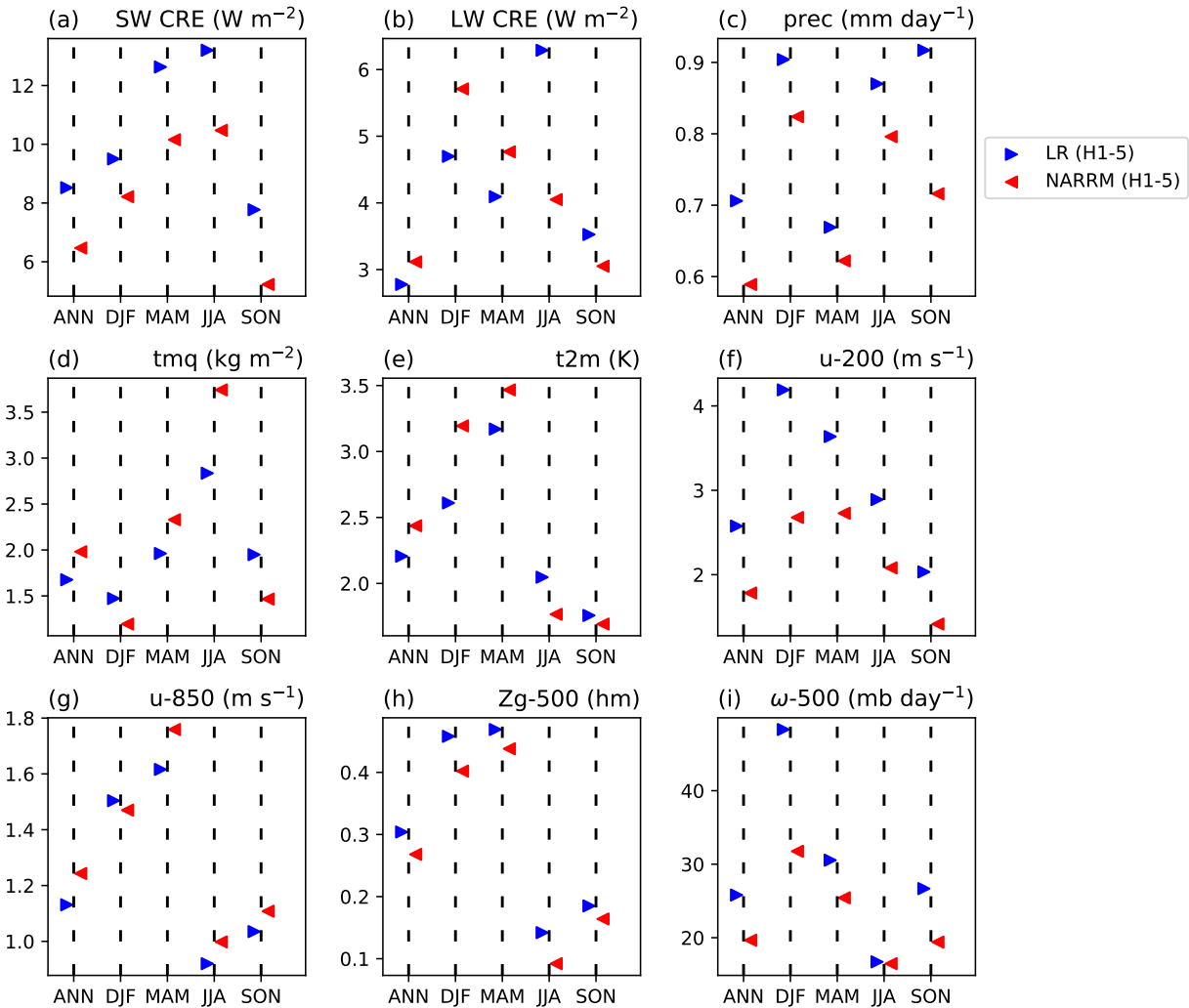
Figure 11. Same as Fig. 7, but contrasting LR and NARRM historical ensemble means at the refined CONUS area (20°– 50°N, 65°– 125°W). Note that variables shown are adjusted to be more appropriate for the CONUS. (a, b) TOA SW and LW cloud radiative effects, (c) precipitation, (d) total precipitable water, (e) surface air temperature, (f, g) 200- and 850-hPa zonal wind, and (h, i) 500-hPa geopotential height and vertical velocity (pressure). The same reference climatology data are used for the variables also shown in Fig. 7, whereas ERA5 (Hersbach et al., 2020) (1985-2014) for (d, e, and i).
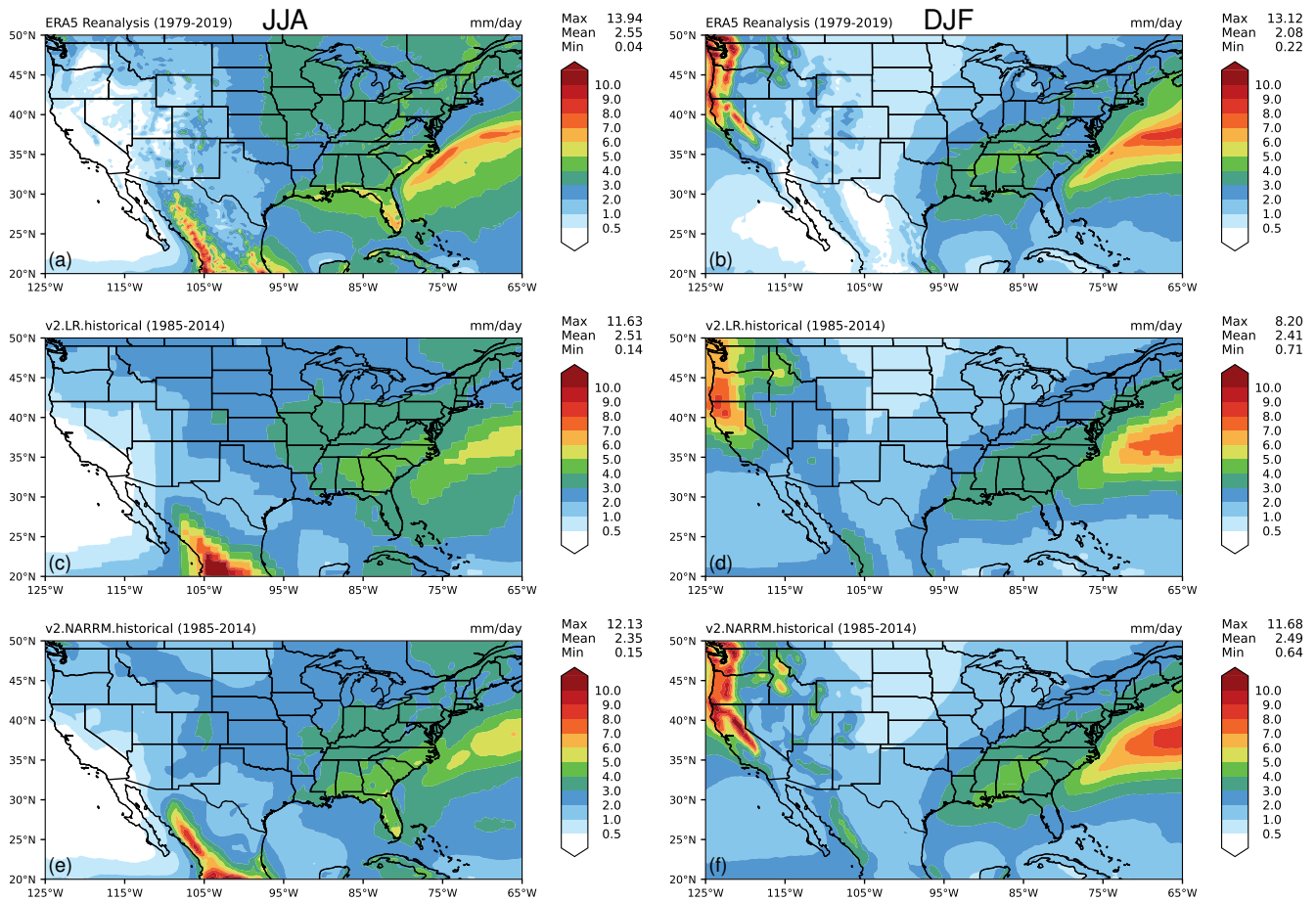
Figure 12. Comparison of CONUS JJA (left) and DJF (right) precipitation geographic patterns from ERA5 reanalysis (a, b), LR (c, d), and NARRM (e, f) historical ensemble means

5.1.2 Aerosols

The E3SMv2 LR model (Golaz et al., 2022) simulates too strong aerosol-related forcing, which has been identified as the primary cause of the underestimated warming in the later portion of historical period in Fig. 9. We will examine here if the

385 NARRM configuration helps bring down the biases in aerosols and anthropogenic forcing by better resolving the meteorological and climate fields. In addition, since NARRM employs the hybrid timestep approach that eliminates re-tuning the scale-dependent aerosol parameters used in LR, e.g., the global scaling factor used to constrain the total emission fluxes of natural aerosols (dust and sea salt) which depend non-linearly on the model-resolved small-scale surface winds, we will also discuss the impact of increasing model horizontal resolution on the natural aerosols and total aerosol optical depth (AOD) in

390 NARRM.
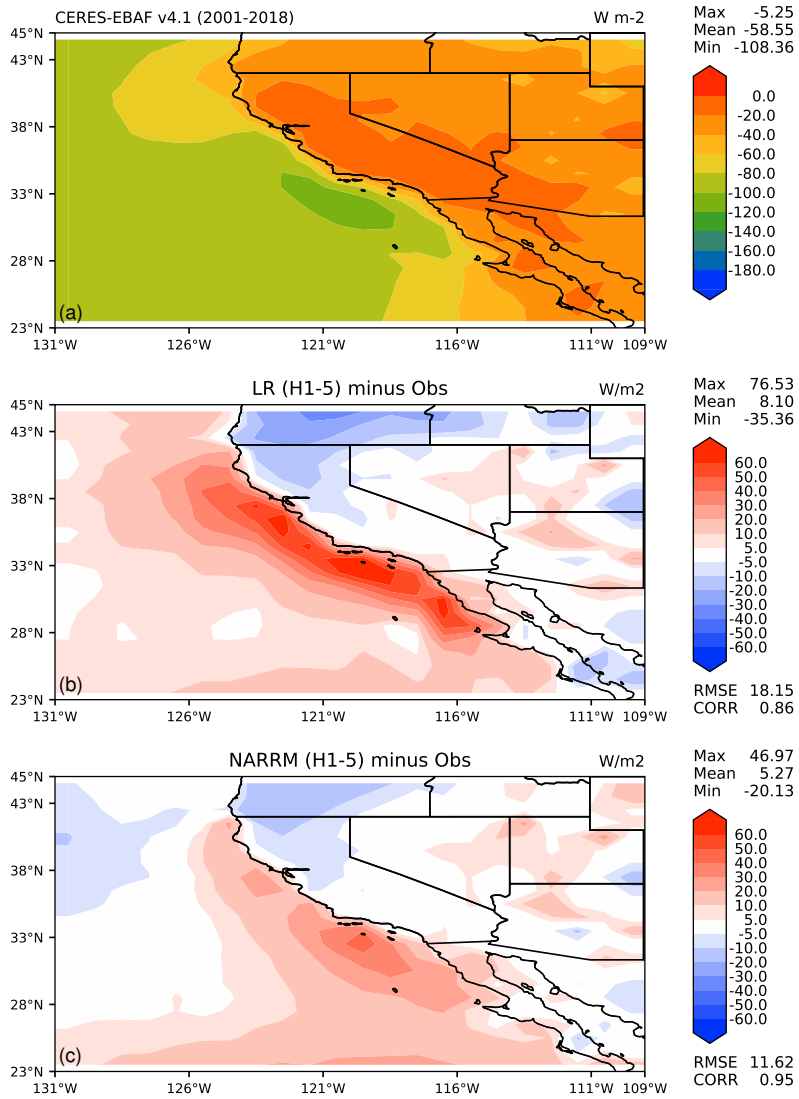
Figure 13. Mean TOA shortwave cloud radiative effects at California in JJA of (a) observations (CERES-EBAF Ed4.1), (b) LR (H1-5) minus observation, and (c) NARRM (H1-5) minus observation
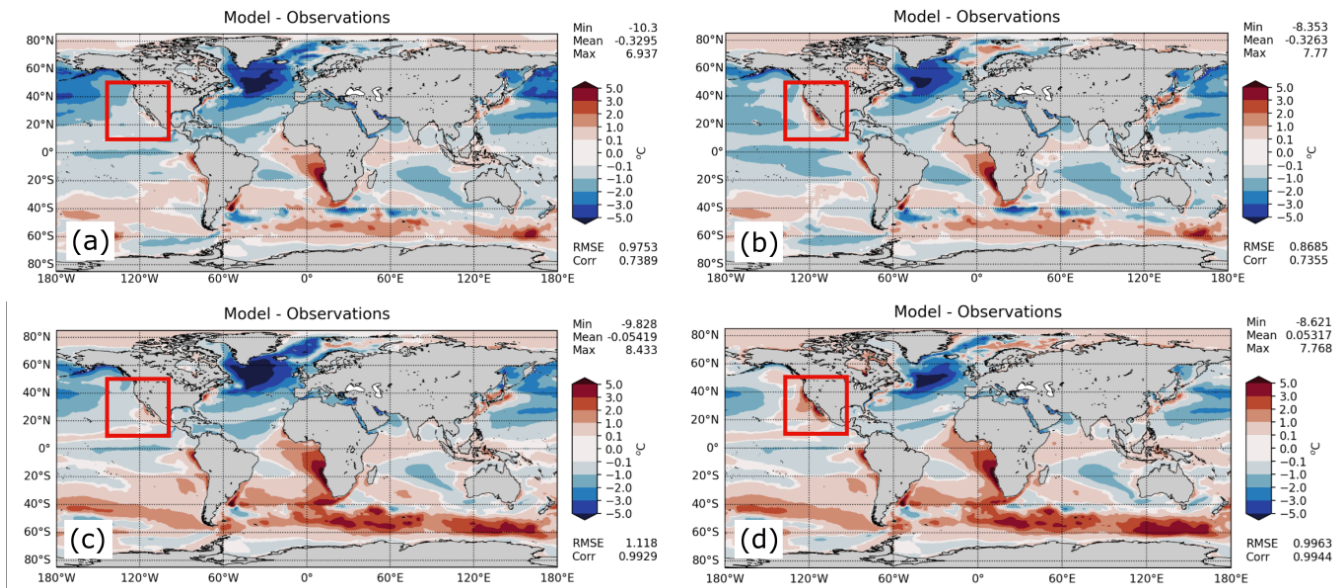
Figure 14. Sea surface temperature (SST) bias (model−observations) simulated by four configurations of E3SMv2 (a) NARRM, (b) LR, (c) RRM atmosphere coupled to LR ocean, and (d) LR atmosphere coupled to RRM ocean. The data is averaged over years 51-100 of the respective piControl simulations.

1. Impact on anthropogenic aerosols

Aerosols in the NARRM configuration are represented in the same manner as in the LR with the enhanced MAM4 (Wang et al., 2020) and improved dust aerosol properties (Feng et al., 2022). Anthropogenic and wildfire emissions used in the LR and NARRM experiments are also from the same input datasets. However, cloud microphysical processes, horizontal advection, and convection can affect aerosol loading within the NARRM high-resolution domain if wet deposition and/or transport are substantially different from the LR configuration (Caldwell et al., 2019). We first compare modeled surface mass concentrations of SO_2, sulfate, black carbon, and organic carbon between the ensemble means of LR and NARRM historical simulations, and evaluate them against ground-based observations from the Clean Air Status and Trends Network (CASTNET) and the Interagency Monitoring of Protected Visual Environments (IMPROVE). The results are shown in Fig. 16. In general, both sets of simulations show a strong correlation with measurements and biases are very similar between the two for the anthropogenic aerosol species. NARRM simulations slightly increase (less than 7%) the concentrations of the four species shown here, compared to LR simulations, which may result from less wet removal or vertical transport in the refined mesh.

2. Impact on natural aerosols

In addition to aerosol removal, emissions of natural aerosols such as dust and sea salt are highly sensitive to resolution changes due to their strong dependence on the resolved surface wind speeds in the model. Increasing model horizontal
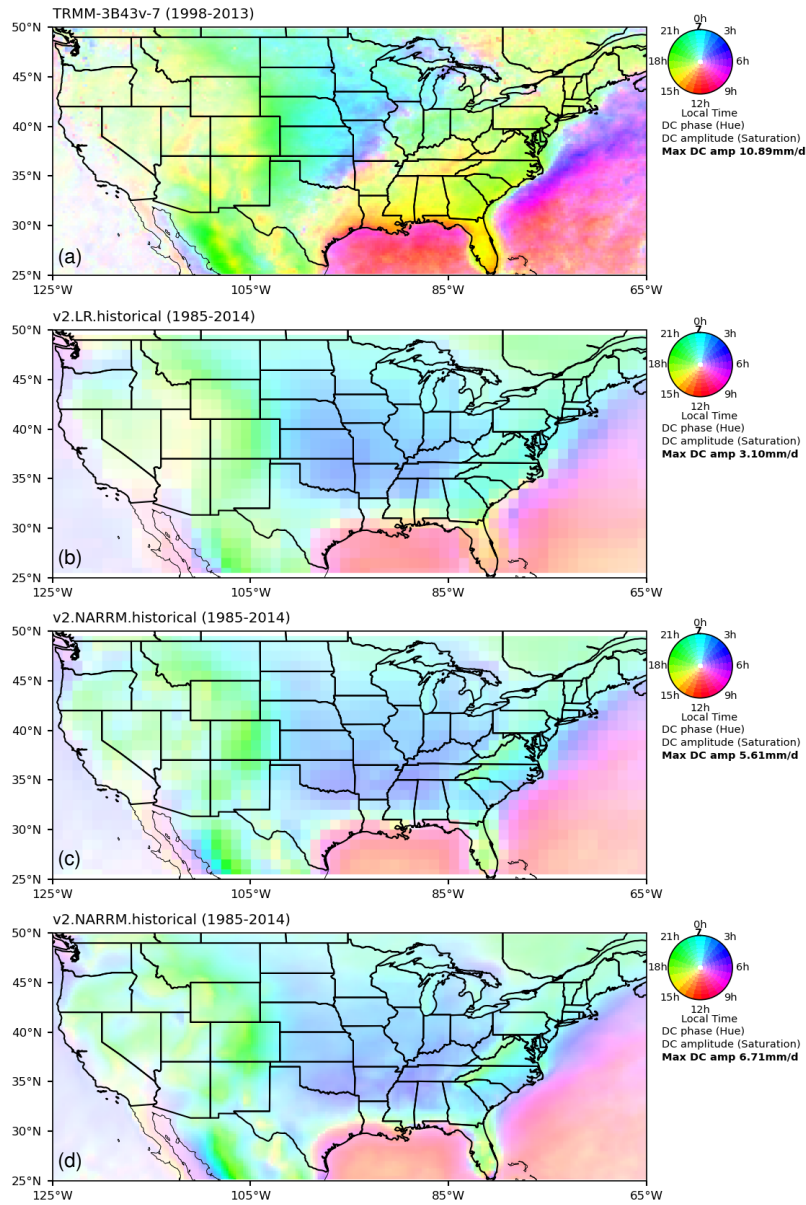
Figure 15. Mean diurnal phase (local time, colors) and magnitude (color density) of the maximum precipitation in JJA calculated from the first harmonic of 3-hourly total precipitation (mm/day) for (a) Tropical Rainfall Measuring Mission (TRMM) observations (Huffman et al., 2007), (b) LR (H1-5), (c) NARRM (H1-5) regridded to the same 100-km grids as (b), and (d) NARRM (H1-5) on 25-km grids.
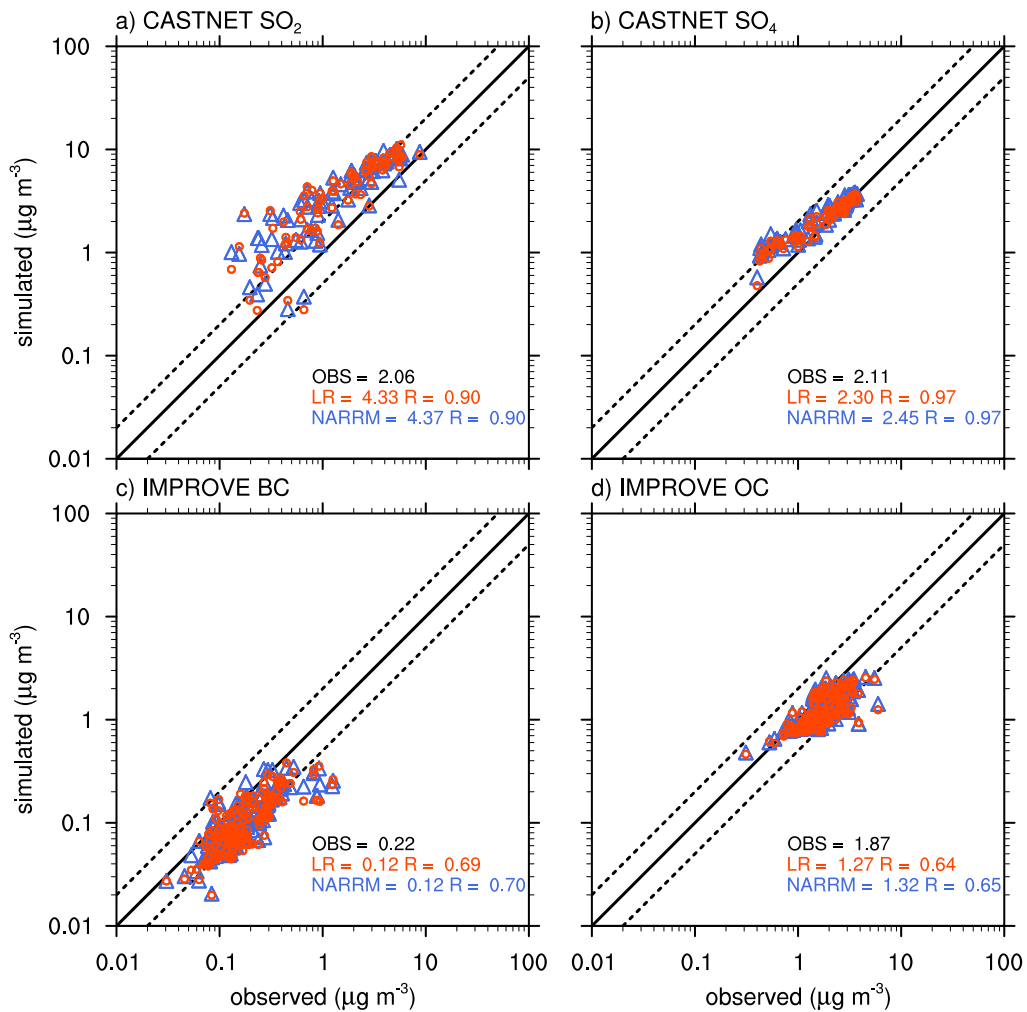
Figure 16. Scatter plots of modeled annual mean surface concentrations of (a) SO₂, (b) sulfate, (c) black carbon, and (d) organic carbon (POM+SOA) from LR (H1-5) and NARRM (H1-5) compared to observations at CASTNET and IMPROVE network surface sites during 2005-2014. The numbers are mean concentration and correlation coefficient (R) for data at the individual sites.

Table 4. Comparison of simulated annual mean AOD (550 nm) between LR (H1-5) and NARRM (H1-5) for the time period of 1985–2014.

AOD	Total	Dust	Sea salt	Sulfate	POM[a]	BC[a]	SOA[a]
			Global means				
LR	0.164	0.028	0.049	0.033	0.009	0.006	0.039
NARRM	0.163	0.028	0.049	0.033	0.009	0.006	0.038
			CONUS means				
LR	0.129	0.0098	0.0205	0.0507	0.0074	0.0057	0.034
NARRM	0.129	0.0101	0.0214	0.0502	0.0075	0.0057	0.034

[a] POM (particulate organic matter), BC (black carbon), and SOA (secondary organic aerosol)

resolution normally requires re-tuning the dust and sea salt aerosol emission factors. For E3SMv1, Feng et al. (2022) showed that without re-tuning, an increase of the horizontal resolution by a factor of four (i.e., from ∼100 km in LR to ∼25 km in HR) results in about 29% increase of global dust emissions and an even larger increase of dust AOD by 42% due to the combined effects from the weakened removal. In contrast, as shown in Table 4, NARRM historical runs simulate nearly the same global mean AODs as LR for all the aerosol species including dust and sea salt, without changing their emission factors. Over the regionally refined CONUS, the mean dust and sea salt AODs are slightly increased (<5%). This suggests that NARRM largely retains the performance of LR for the aerosol simulations on the global and regional mean basis, without requiring additional re-tuning of the scale-dependent emission factors.

3. Aerosol spatial variability and extremes

On the other hand, NARRM shows improvement over LR in representing aerosol spatial variability and extreme values over the refined mesh region. Figure 17 compares the simulated AOD (550 nm) distributions between LR (0101) and NARRM (0101) historical simulations for the present-day time period of 2000–2014. While both depict a similar general geographical pattern, e.g., higher AODs over the more polluted eastern US than the western part of the country, NARRM captures greater and finer detail in spatial variability than LR, e.g., over the mountainous areas along the Rockies, Sierra Nevada, and Appalachians. The better resolved AOD variability in NARRM results from the spatial refinement of the resolution-dependent aerosol emission fluxes (natural species), transport, and removal as discussed above. Compared to the ground-based AOD measurements at the 37 AERONET (Holben et al., 1998) sites (2006–2015), NARRM shows stronger spatial correlation with the observations than LR (Fig. 17c). Both configurations overestimate the mean AOD averaged over the AERONET sites, possibly linked to the weak wet removal in E3SMv2 (Golaz et al., 2022).

In addition to the improved spatial variability, higher resolution in NARRM also leads to more frequent occurrences of large AOD predictions over CONUS than the LR model, especially over the regions dominated by wind-driven dust or sea salt aerosols. Figure 18 shows an example of the calculated probability density function (PDF) for dust AOD over the major dust source region in the US (32–42°N, 118–108°W; indicated by the 10°x10° box in Fig. 17b), from both the LR (0101) and NARRM (0101) simulations in 2000–2014, which are remapped to the same 0.25° grid resolution.

(a) LR (0101)

(b) NARRM (0101)

(c) Comparison with the AERONET AOD
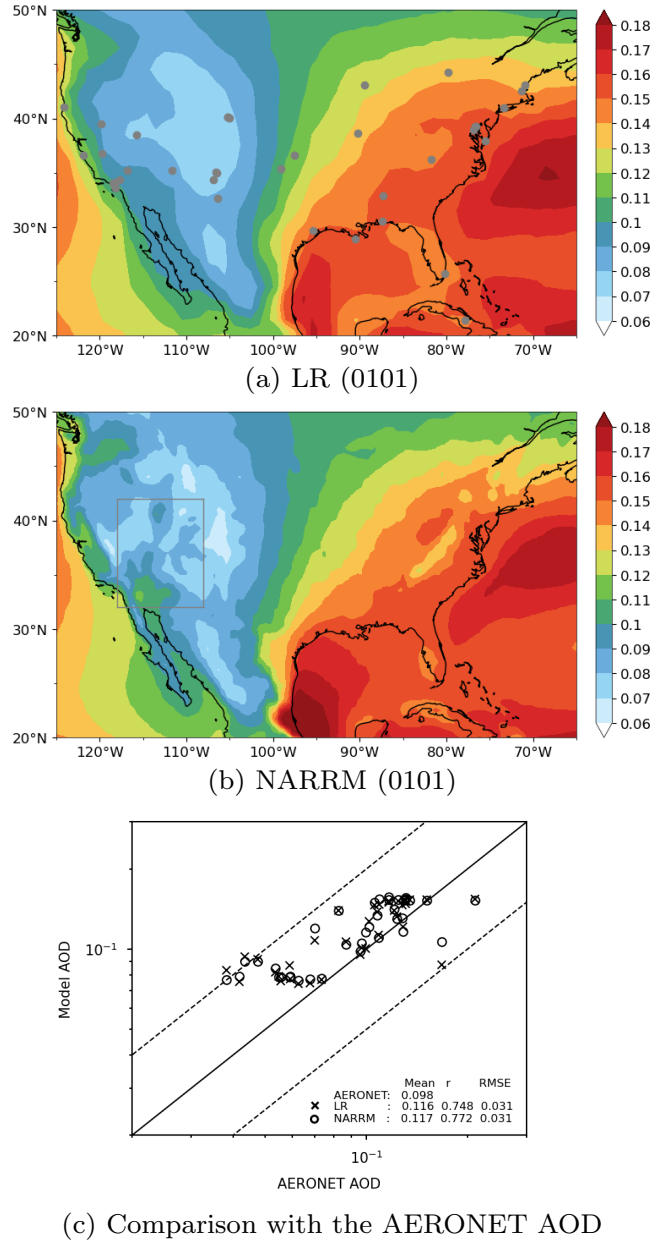
Figure 17. Aerosol optical depth (AOD) at 550 nm from (a) LR (0101) and (b) NARRM (0101) historical simulations averaged over 2000–2014. Panel (c) shows the AOD comparison of the two model simulations with the AERONET observations during 2006–2015. The site locations of AERONET are denoted by the gray dots in panel (a). The gray box in panel (b) denotes the dust region referenced in Fig. 18.

It is worth noting that the remapping of the LR results to the finer resolution leads to little improvement in the resolved spatial variability in dust AOD. Clearly, NARRM predicts more occurrences of high dust AOD over this region than LR, e.g., 22% of the dust AODs predicted by NARRM exceeds 0.015, which is the top 98th percentile of the LR model predictions remapped to the same resolution. This suggests that LR may significantly underestimate the occurrences of large dust outbreaks in the southwestern US region relative to NARRM, due to the unresolved surface winds for dust mobilization in the model. Similarly, NARRM would be more suitable for urban climate or air quality studies for capturing the extremely polluted cases occurring at finer spatial or temporal scales.
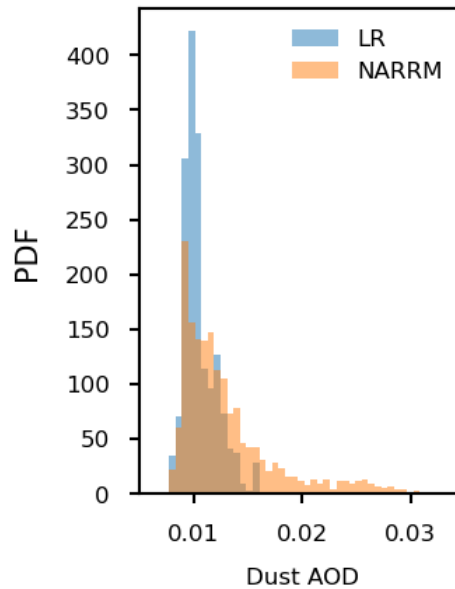
Figure 18. Calculated probability density function (PDF) of the dust AOD predictions from LR (0101) and NARRM (0101) between 2000–2014 remapped to the same 0.25° grid resolution, over the major dust source region in the US (32–42°N, 118–108°W; indicated by the 10°×10° box in Fig. 17b).

4. Effective radiative forcing of anthropogenic aerosols

Figure 19 shows the effective radiative forcing of anthropogenic aerosols (ΔF) over CONUS and adjacent ocean areas estimated using nudged LR and NARRM simulations. ΔF is overall negative in both LR and NARRM and dominated by the shortwave component (ΔF_{SW} in Figure 19be). The regional mean ΔF_{SW} and ΔF_{LW} are both slightly stronger (more negative for ΔF_{SW} and more positive for ΔF_{LW}) in NARRM compared to LR. Over the Pacific ocean near 20°N and 120°W, ΔF_{SW} and ΔF in NARRM are much stronger than in LR. This is mainly caused by larger low cloud fraction simulated in NARRM (see Appendix Fig. A4), which causes a larger contrast in droplet number concentration and liquid water path between the PD and PI simulations compared to LR (Fig. A5).
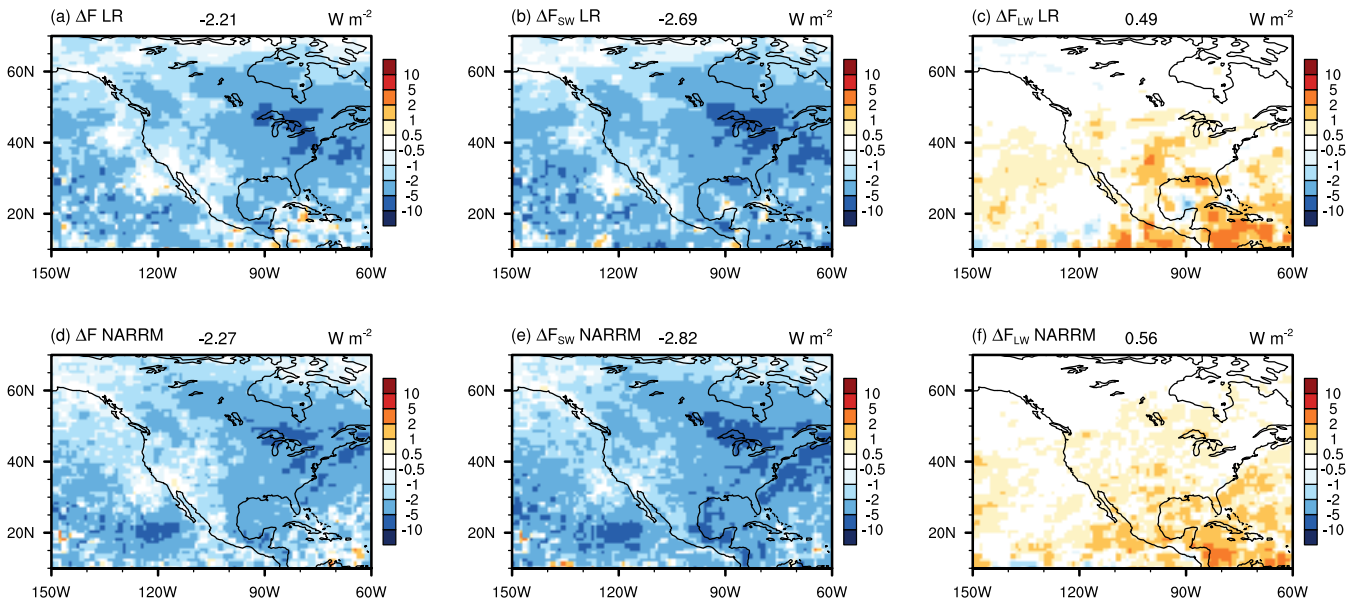
Figure 19. Anthropogenic aerosol effects simulated by the nudged LR and NARRM simulations.

5.1.3 Cloud and cloud feedback

Here we examine the impact of increased horizontal resolution over NA on the simulated clouds and their radiative effects with the LR and NARRM historical simulations and on cloud feedback changes with the quadrupling $4xCO_2$ simulations.

Figure 20 compares the cloud cover between the E3SM CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) simulator output and the GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP) (Zhang and et al., 2023). Cloud cover and cloud thermodynamic phase are diagnosed with the same algorithm in the CALIPSO simulator and CALIPSO-GOCCP data, facilitating consistent model-observation comparisons (Chepfer et al., 2008, 2010; Cesana and Chepfer, 2013). Observed total cloud cover is larger over the NA polar region than the CONUS in the CALIPSO-GOCCP data. Total cloud cover is larger than 60% over the eastern Pacific Ocean, North Atlantic Ocean, and Arctic Ocean. Strong land-ocean contrast is observed — liquid phase clouds dominate over the ocean while ice phase clouds prevail over the land such as Greenland and CONUS. Compared to CALIPSO-GOCCP, LR overestimates total cloud cover at NA high latitudes and western CONUS, and underestimates in Greenland, particularly over the Baffin Bay and near the Greenland coast. The underestimated cloud cover to the west coasts of CONUS is also notable, which is consistent with the previous discussion on Fig. 13. The excessive modeled total cloud cover is primarily attributed to the positive biases in the liquid cloud over the polar region. The positive biases of ice cloud cover contributes to the biases over the mountainous regions in the western NA. On the other hand, ice clouds are underestimated over Greenland and the northern Canada.

Over land, NARRM displays improvements relative to LR in the western NA and Arctic for both cloud phases. For instance, the ice cloud biases are significantly reduced from Alaska to the western CONUS (Fig. 20fi), and the liquid cloud deficiencies

over Alaska and Greenland are generally decreased (Fig. 20eh). The better represented topography in NARRM (Fig. 8) is probably the key factor of these NARRM improvements. The impact of increased horizontal resolution on simulated cloud phase is also noted in the E3SMv1 model with the CALIPSO simulator (Zhang et al., 2019), where increased horizontal resolution also slightly decreases simulated liquid and ice clouds at temperatures warmer than $-40°C$ in the Arctic region.

Over ocean, NARRM substantially improves the stratocumulus clouds to the west of coastal regions. NARRM also moderately outperforms LR in representing liquid clouds over the North Atlantic to the west of Greenland. This is related to the warmer ($\sim1.5°C$) NARRM surface air temperature over the Labrador Sea. This warmer NARRM surface air temperature is consistent with the decreased sea ice concentration in that region (not shown), which is somewhat expected as an advantage of refining grids for both atmosphere and ocean/sea ice (see Fig. A1). Further process-level analysis is necessary to fully understand this LR-NARRM model behavior change and will be reported in separate papers.

Figure 21 compares the simulated SW and LW CRE with the CERES-EBAF Ed4.1 observations. Large negative SW CRE biases and positive LW CRE biases are shown over the Arctic land and west coasts of NA (i.e., Alaska to Oregon) in LR, which mainly result from the overestimated cloud cover. These biases are substantially reduced in NARRM primarily owing to the improved cloud cover (Fig. 20). Given the reduced negative bias of marine stratocumulus clouds in NARRM near the west coasts of CONUS, simulated SW CRE is also largely improved. As discussed by Golaz et al. (2022), sea ice concentration is largely overestimated over the North Atlantic Ocean in LR. The too large sea ice extent leads to weaker SW and LW CRE than observed at Labrador Sea. This is primarily because of the brighter and colder sea ice surface in LR that reflects more SW radiative fluxes and emits less LW radiative fluxes than the observations under clear-sky conditions (not shown). Compared to LR, the maximum positive bias in NARRM sea ice extent in Labrador Sea is greatly alleviated. The better simulated sea ice extent reduces the biases of overly reflective clear-sky SW radiation and the insufficient clear-sky outgoing LW radiation. With generally comparable all-sky SW and LW radiative fluxes between LR and NARRM, those reduced clear-sky biases thus leads to a better CRE in NARRM over Labrador Sea.

Given the improved historical cloud cover and cloud radiative effects over the NA, we further examine the regional climate feedbacks over this region in Fig. 22. Relative to the global-mean value, the total climate feedback over NA is more negative from NARRM than from LR. This mainly results from the more negative Planck feedback and less positive SW cloud feedbacks. The more negative Planck feedback is related to the stronger surface warming over the Northeastern Pacific (not shown).

Figure 23 shows the spatial distribution of cloud feedback of LR and NARRM in the NA region. Due to the different cloud types over land and ocean, we report their regional averages separately. Notably, the total land cloud feedback is 0.41 W/m^2/K smaller in NARRM (0.26 W/m^2/K) than in LR (0.67 W/m^2/K), which is dominated by the reduced SW cloud feedback over the Northeastern US (Fig. 23e). Further examination indicates this reduction is mainly related to the weaker reduction in low cloud cover under warming in NARRM. Figure 20g presents that the overestimated cloud cover is slightly alleviated over the Arctic land region in NARRM, implying that the less mean-state cloud cover might contribute to a weaker cloud reduction under warming there. Over ocean, NARRM presents a stronger SW cloud feedabck and a weaker LW cloud feedback over the marine low cloud regime, leading to a small change of total cloud feedback. Across the CSS/WGNE Pacific Cross-Section

Intercomparison (GPCI) transect (Teixeira et al., 2011), NARRM tends to show a weaker positive cloud feedback near the coast and more positive cloud feedback off the coast. These suggest that the regional refinement can significantly affect the regional cloud responses under warming and the predictability of regional climate.

5.1.4 Extratropical cyclone

One of the primary motivations for pushing climate simulation resolution is to potentially better capture extremes. Extratropical cyclones (ETC) are a major weather extreme phenomenon in middle and high latitudes, bringing with them strong winds and precipitation that can exert substantial societal impacts along their pathways over days and over hundreds to thousands of kilometers. Climate changes are likely to induce changes to the dynamical and physical characteristics of ETCs as well as their geospatial distribution (e.g. Bengtsson et al., 2006; Ulbrich et al., 2009). Projections of such future changes rely highly on numerical climate and ESM models. Their skills in simulating major weather systems like ETC have been carefully scrutinized by modeling centers and by the climate science community in conjunction with the major intercomparison campaigns (Greeves et al., 2007; Chang, 2013). While the conventional climate models with grid resolution around 100 km show reasonable skills in producing ETC frequency and spatial track density, it has also been found that higher resolution models are better capable of capturing more intense ETCs (e.g. Jung et al., 2006), which is critical for using ESM to project future climates as growing evidence shows that global warming tends to shift the weather spectrum to the more extreme end (Melillo et al., 2014). Here, we will demonstrate the benefits of higher resolution in simulating ETCs in a regionally refined setting, by comparing the results from NARRM with those from LR simulations against the ETC activities derived from the ERA5 reanalysis.

The ETC tracks and statistics can be obtained using automated identification and tracking algorithms (e.g. Blender and Schubert, 2000; Geng and Sugi, 2001; Bengtsson et al., 2006; Jung et al., 2006; Ullrich and Zarzycki, 2017; Ullrich et al., 2021). The objective identification and tracking also make it suitable to compare ETC activities and statistics derived from different data sources in particular for model evaluations. The algorithms usually identify and track the spatial features of a meteorological variable, such as mean sea-level pressure (MSLP) or 850 hPa vorticity, that can characterize the structure of cyclones and their movements. In this work, we use a community feature detection and tracking framework, TempestExtremes (Ullrich and Zarzycki, 2017; Ullrich et al., 2021) to derive ETC activities from 6-hourly MSLP data during the period of 1985–2014 from the E3SM simulations and the ERA5 reanalysis. Considering higher-resolution data can more accurately identify the storms and their tracks (Blender and Schubert, 2000; Geng and Sugi, 2001), all the model and reanalysis data are placed on 1°x1° grids to feed the tracking software. The algorithm takes two steps. First, a candidate cyclone is detected when a minimum MSLP feature is enclosed by a contour of 200 Pa interval within 6 degrees of the center. Candidates within 6 degrees of one another are merged, with the lower center pressure taking precedence. The candidates are then stitched together to define the tracks if the features persist for at least 60 hours with a maximum gap of at most 18 hours. From the start to the end, a candidate cyclone must travel at least 12-degree great circle distance to qualify as an ETC.

Over the NARRM high-resolution domain, ETCs are most active during winter. Figure 24 shows the mean DJF track density for the models and the analysis derived by casting the computed ETC track data onto 5°x5° grids. The tracks are mostly concentrated over the Northeast Pacific and Northwest Atlantic that form the well-known storm tracks. Both LR and NARRM
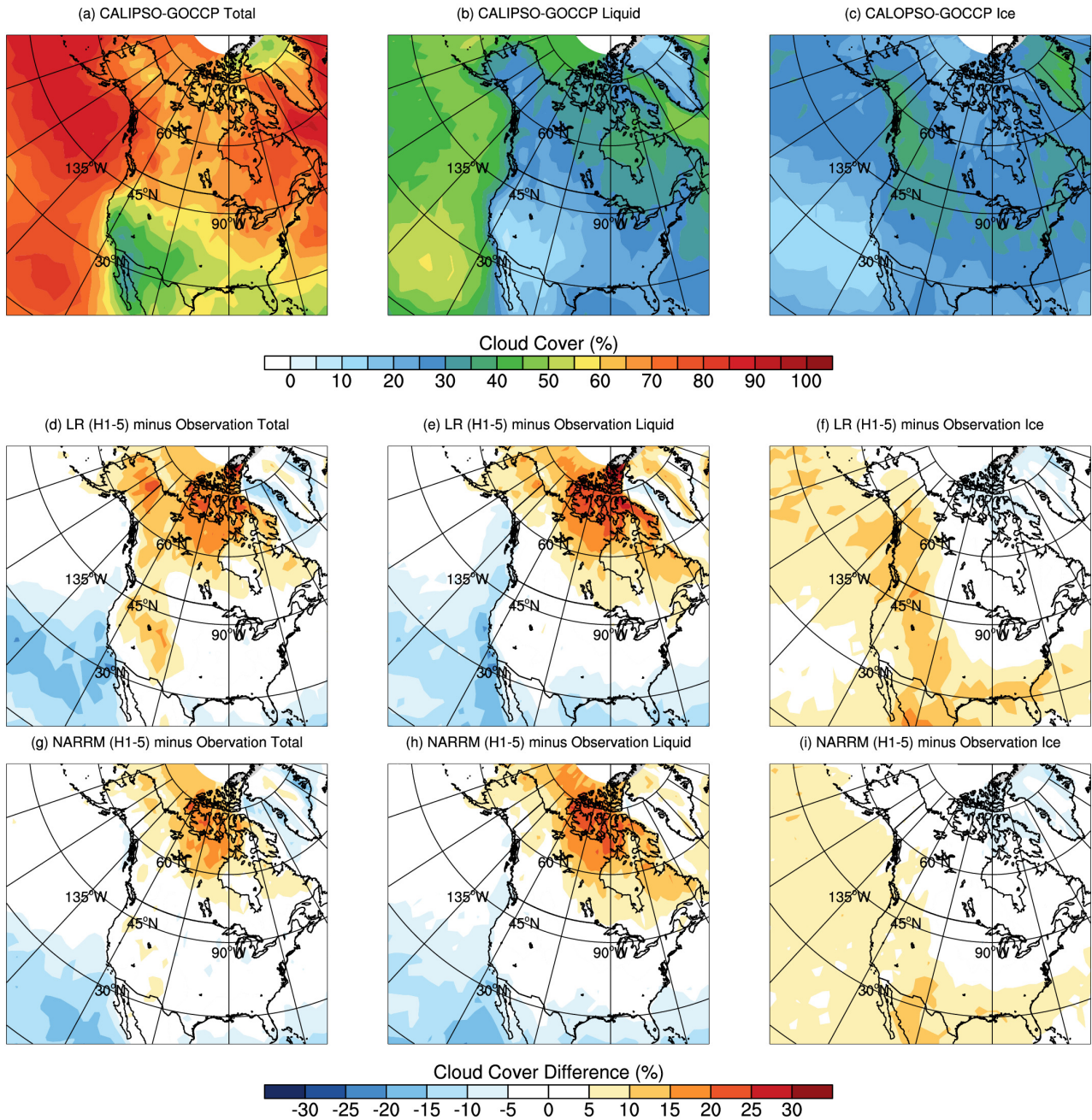
Figure 20. Spatial distribution of annual mean total cloud cover (a), cloud cover in liquid phase (b), and ice phase (c) from the CALIPSO-GOCCP data. The cloud cover biases in LR (H1-5) and NARRM (H1-5) historical simulations (1985–2014) are shown in (d–f) and (g–i), respectively. Simulated cloud cover and cloud thermodynamic phase are derived by the CALIPSO simulator. Climatology data of CALIPSO-GOCCP version 3.1.2 (Chepfer et al., 2010) from 2006–2018 are used in the model evaluation.
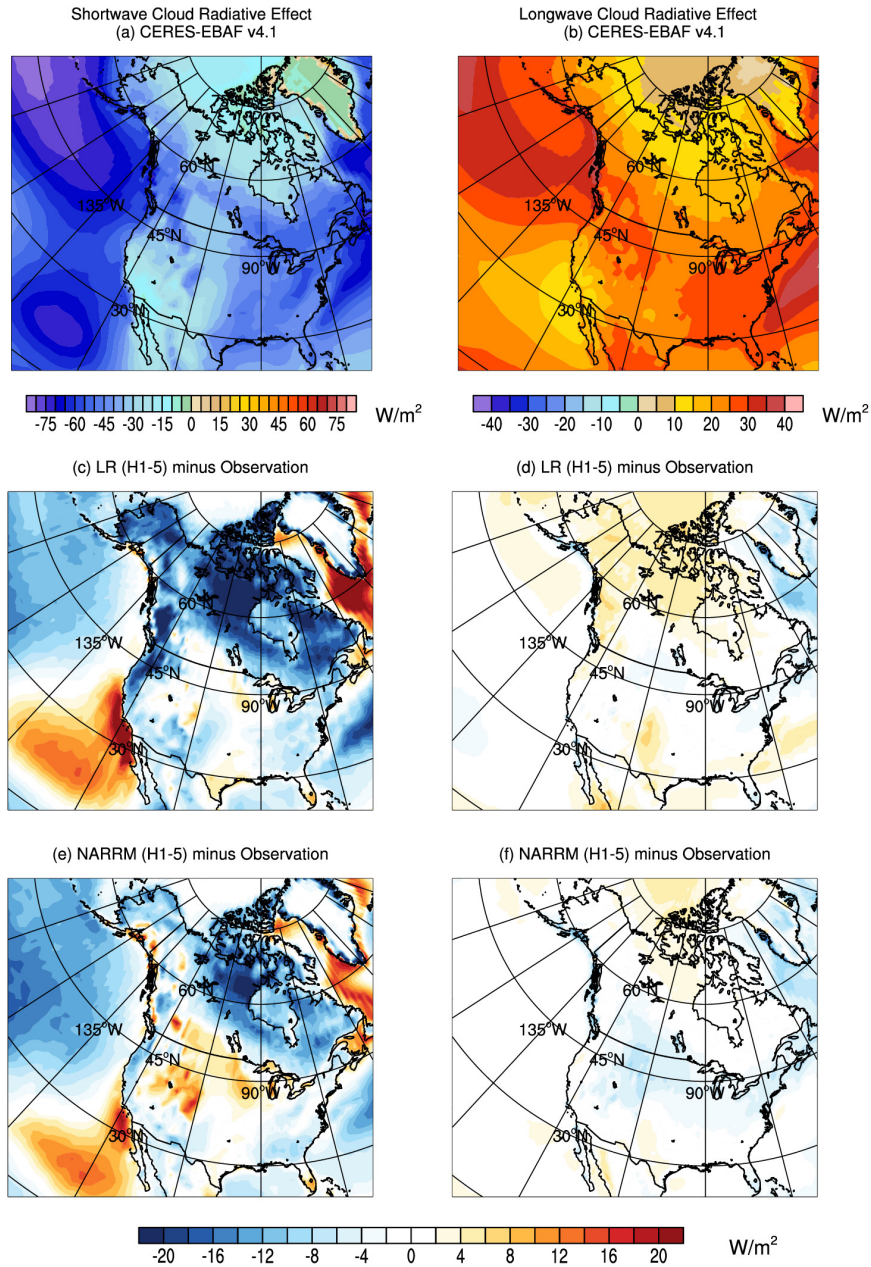
Figure 21. Spatial distribution of observed shortwave cloud radiative effect (a), longwave cloud radiative effect (b), and simulated cloud radiative effect biases in LR (H1-5) (c, d) and NARRM (H1-5) (e, f) historical simulations (1985–2014). The observed cloud radiative effect is from the CERES-EBAF Ed4.1.
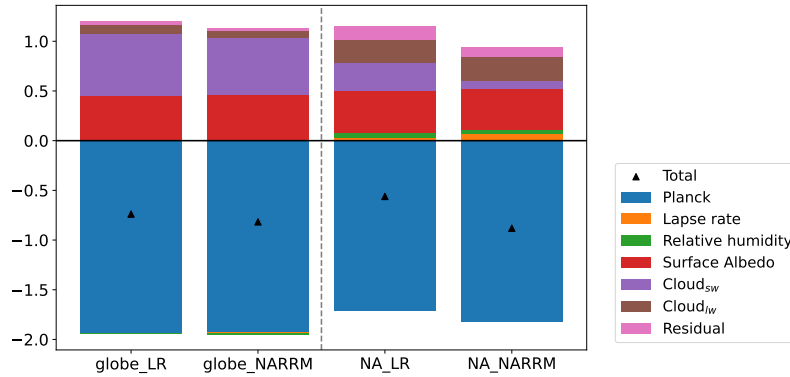
Figure 22. Global- and NA-mean climate feedbacks of LR and NARRM decomposed using radiative kernels (e.g. Soden et al., 2008; Held and Shell, 2012).

simulations capture these main features to a large extent. There are, however, notable differences between LR and NARRM over these oceanic storm tracks. The track densities are clearly underestimated in the LR simulations inside the refined region, except for the Atlantic storm track in the coupled mode. NARRM clearly produces higher ETC track density than does LR for both sections of the oceanic storm tracks, and mostly agrees better with the ERA5 data, although in the coupled mode, the track density tends to be overestimated. The shape and orientation of both the Pacific and Atlantic storm tracks are much better produced by NARRM in the coupled mode. Given the significant differences from both coupled LR (Fig. 24d) and uncoupled NARRM (Fig. 24c), it is reasonable to believe that the better captured storm track shapes in the coupled NARRM are due to interactions with the refined ocean (Fig. A1). Several secondary centers of active ETCs in the ERA5 over land are also reproduced by the models, including the active regions over the Great Lakes and the Hudson Bay, though the densities are overestimated in the models, more so in the NARRM. It is worthwhile to mention that the NARRM simulations are able to produce the chain of secondary centers to the east of the Rocky Mountains that are also present in the ERA5 reanalysis but are largely missing in the LR simulations. This is presumably due to the NARRM's better resolved mountainous terrain features, a benefit by design.

The benefit of grid refinement can be further seen in Fig. 25, which shows the histograms of the ETC as a function of the minimum center pressure and the maximum deepening rate during its lifetime. All events within the refined region bounded by the black dashed lines as shown in Fig. 24 are used to compute these statistics. The maximum deepening rate is defined as the maximum 6-hourly center pressure drop, normalized by $\sin(\varphi_{\text{ref}})/\sin(\varphi)$, with φ being the latitude and φ_{ref} the reference latitude at 45 degrees (see also in Jung et al., 2006). Clearly the NARRM very closely reproduces the number of intense cyclones (minimum center MSLP < 960 hPa), while not surprisingly the LR model underproduces. These are true in coupled and uncoupled modes. Both LR and NARRM simulations overestimate the number of weaker ETCs. On a similar note, the observed number of rapid-growth cyclones are closely reproduced by the NARRM simulations, while clearly underestimated by a large margin in the LR simulations.
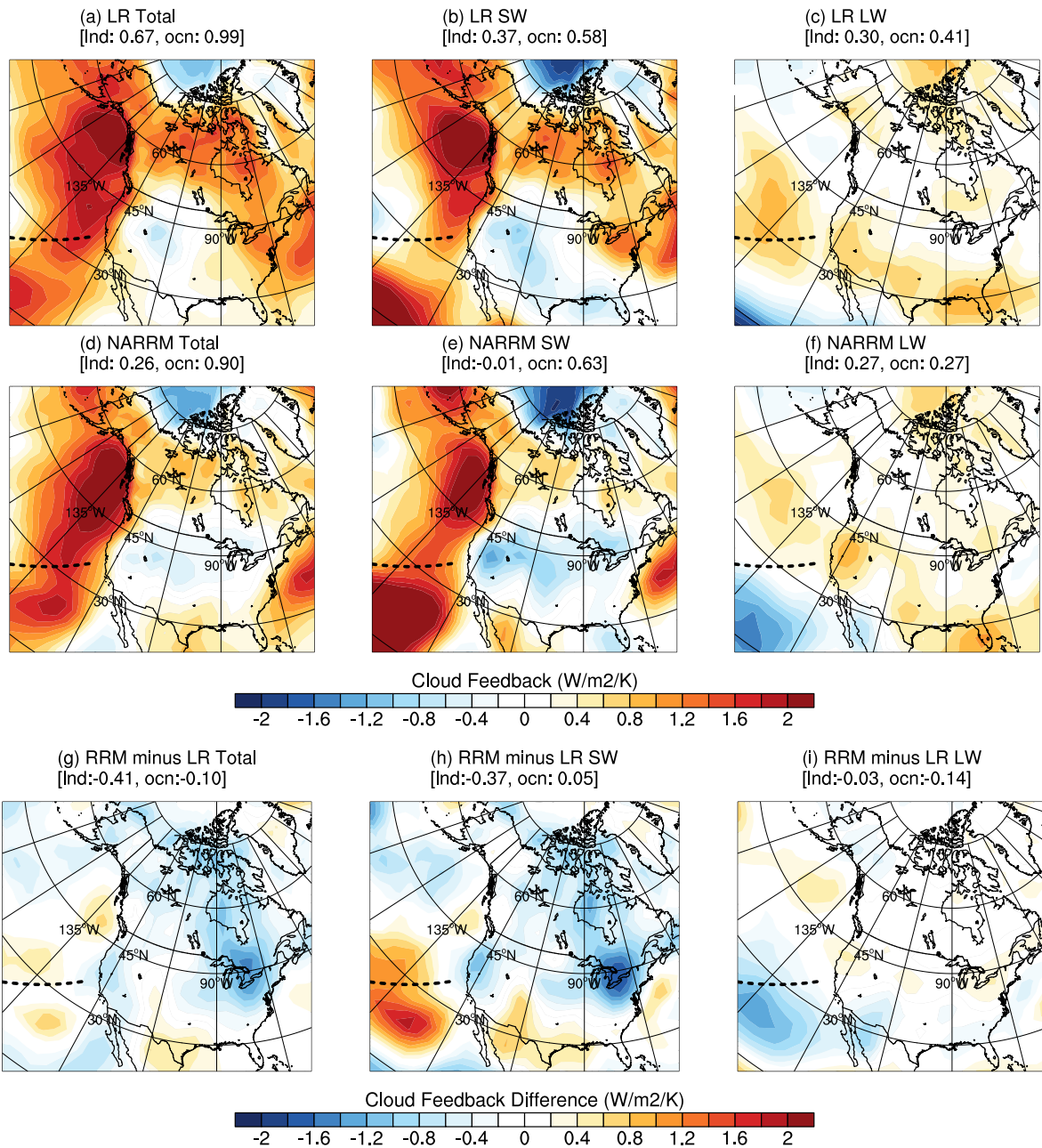
Figure 23. Spatial distribution of North America (NA) total (a, d, g), SW (b, e, h), and LW (c, f, i) cloud feedbacks for LR (a-c), NARRM (d-f), and the difference between NARRM and LR (g-i). The GPCI transect is denoted by the black dashed line. The average values over land and ocean are labelled in the brackets.
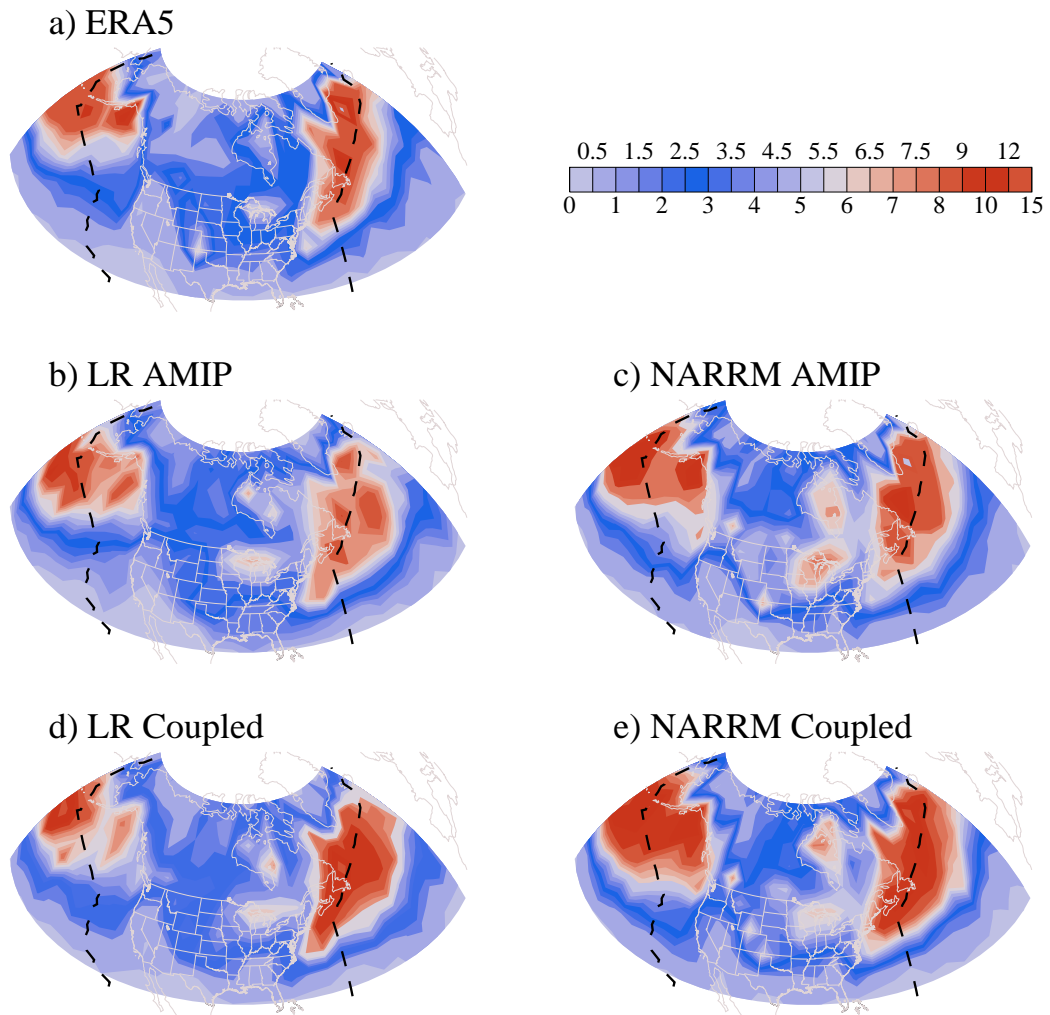
Figure 24. Mean extratropical cyclone track density in the DJF season between 1985 and 2014 from (a) ERA5, (b, c) AMIP, and (d, e) historical simulations of LR and NARRM. Black dashed lines denote the western and eastern boundary of the refined region.
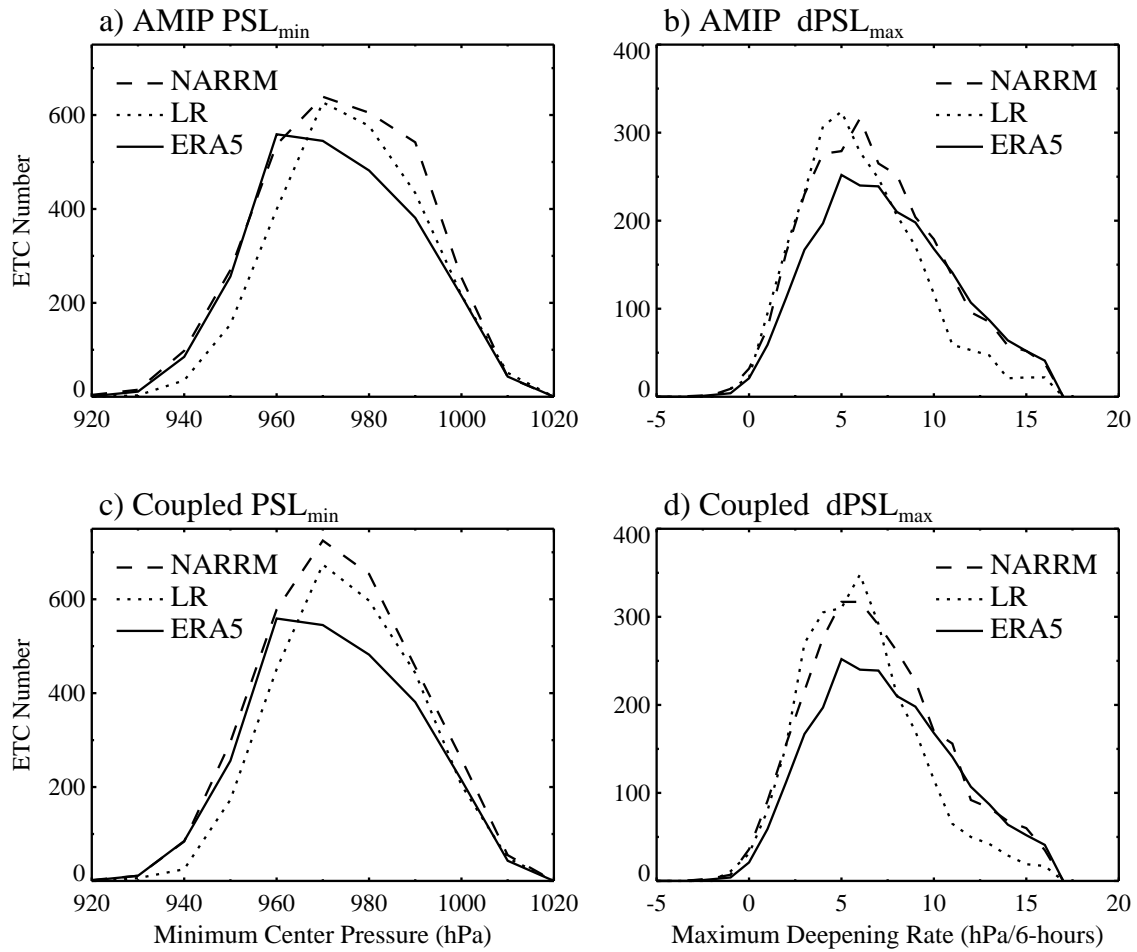
Figure 25. Histogram of extratropical cyclone minimum center sea level pressure and maximum 6-hour deepening rate in the DJF season between 1985 and 2014 (a, b) AMIP and (c, d) historical simulations. LR is shown in dotted, NARRM in dashed, and ERA5 in solid black lines.

 5.2   Land and river

5.2.1   Snowpack

Natural storage provided by mountain snowpack is central to water supply reliability in the western U.S. (Siirila-Woodburn et al., 2021). To evaluate model skill in representing this critical hydroclimate benchmark variable, intra-annual snowpack dynamics are evaluated using the methodology known as the snow water equivalent (SWE) triangle (Rhoades et al., 2018a, b).
560   The seven metrics that make up the SWE triangle attempt to distill management-relevant aspects of the accumulation and ablation of snowpack (e.g., peak water volume and snowmelt rate) for any arbitrary gridded SWE dataset. Five HUC2 basins of the mountainous western U.S. are used to derive five-member ensemble- and basin-average evaluations of LR and NARRM fully-coupled historical simulations and are compared with ERA5. All datasets are bi-linearly regridded using ESMF to 0.25° resolution prior to masking and computing the basin-average SWE triangle metrics.
565   NARRM provides enhanced winter (DJF) climatological representation of the spatial variability of SWE across the CONUS relative to LR (Fig. 26ab). This is seen through higher SWE magnitudes and more granular spatial structures in NARRM compared with LR, particularly in coastal mountain ranges such as the Cascades and Sierra Nevada, and corroborates a long-history of ESM studies that highlight the critical importance of horizontal resolution ($\leq$0.25°) in properly representing the mountainous hydrologic cycle (Demory et al., 2014; Rhoades et al., 2017; Kapnick et al., 2018; Palazzi et al., 2019; Bambach
570   et al., 2021; Rhoades et al., 2022). As shown through the more granular intra-seasonal perspective of the SWE triangle metrics, certain aspects in the snowpack dynamics are improved with NARRM (e.g., peak water volume), namely in the Pacific Northwest and California (Fig. 26c). With that said, some E3SM SWE biases are not ameliorated with horizontal resolution and may arise due to the combination of higher winter-season precipitation (Fig. 12) and general cool-bias (Fig. 9) in both the LR and NARRM fully-coupled historical simulations.

575   **5.2.2   Runoff and evapotranspiration**

NARRM better captures spatial variability in land hydrologic processes, as indicated by two most important land hydrologic variables, total runoff (Fig. 27) and evapotranspiration (ET) (Fig. 28). For instance, in the pacific coastal regions, the Rocky Mountains block atmospheric moisture from ocean to inland areas and lead to two distinct hydrologic regimes: a wet regime in the mountain west and a dry regime in the mountain east. This abrupt spatial shift from a wet to dry hydrologic regime can be
580   clearly seen in the composite runoff map from the Global Runoff Data Center (GRDC) (Fekete et al., 2011), Fig. 27a, and the observed evapotranspiration map from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite observations (Running et al., 2017), Fig. 28a. Note that the GRDC runoff map is not completely based on the observational data since runoff measurements are not available at the regional or global scales due to technical and economic limitations. It is nevertheless a more realistic estimate than any model simulations because it was first generated with a monthly hydrologic model (hence
585   producing spatiotemporal variability) and then bias-corrected against discharge measurements at thousands of river gauges (Fekete et al., 2011). This abrupt shift of hydrologic regime around the Rocky Mountains, along with the other spatial variations, is much better resolved in the NARRM simulation than LR, which is the case for both simulated runoff and evapotranspiration,
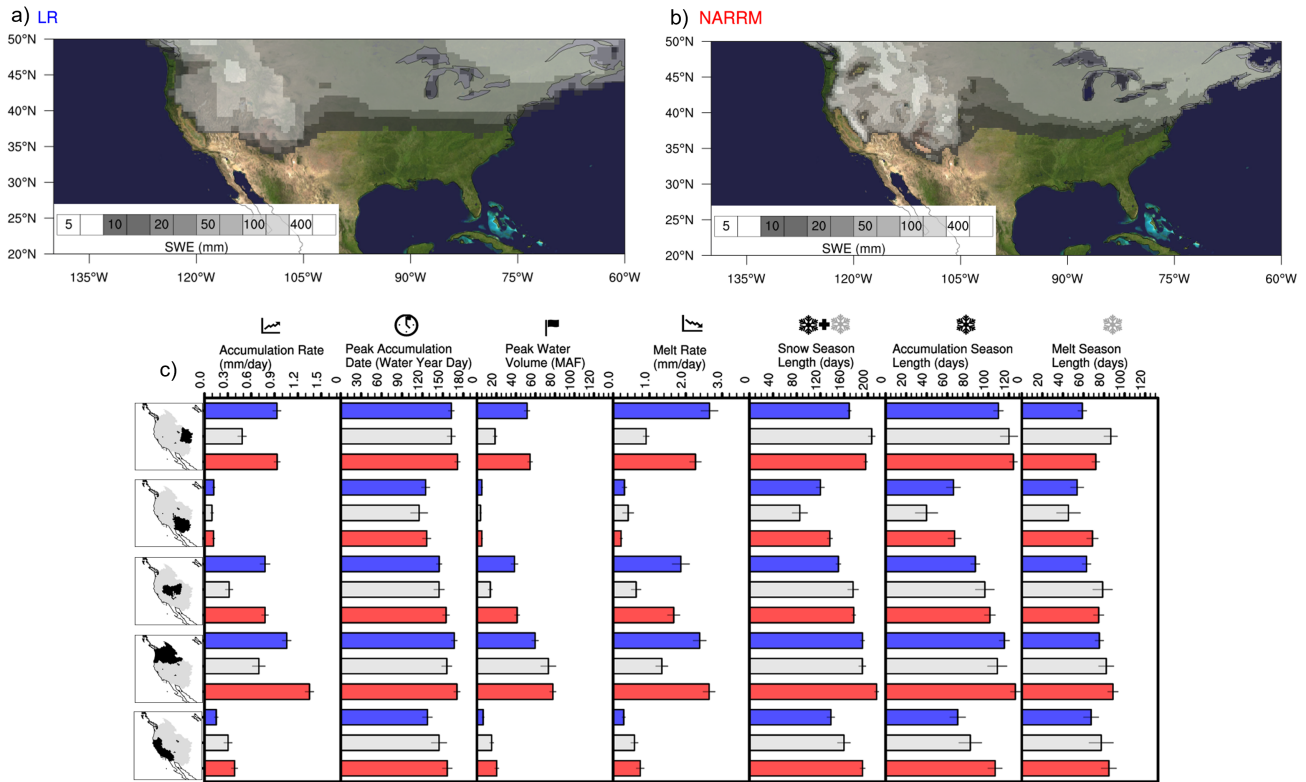
Figure 26. Climatological DJF average snow water equivalent (SWE) as simulated by E3SMv2 with (a) LR and (b) NARRM over the 1985–2014 period. (c) LR (blue) and NARRM (red) SWE triangle metrics for five HUC2 basins within the mountainous western U.S. compared with ERA5 (gray). Black-bars at the end of each histogram represent the mean 95% confidence intervals.

as shown in Figs. 27bc and 28bc. The NARRM simulated spatial patterns are thus more realistic than the LR ones over NA. Over the remaining regions of the globe, the NARRM and LR simulated spatial patterns are quite similar to each other in terms

590   of both runoff and evapotranspiration (not shown).

The simulation biases in runoff and evapotranspiration are further examined in terms of absolute biases, i.e., absolute difference between the simulated and "benchmark" values. Here the GRDC runoff and MODIS ET data are used as the "benchmark" data. Figures 27d and 28d show the maps of absolute bias difference, i.e., difference between the absolute biases in the LR simulation and those in the NARRM simulation (former subtracting latter), for annual mean runoff and evapotranspiration,

595   respectively. For a specific grid cell in these two maps, a positive difference means the absolute bias in the LR simulation is larger than that in the NARRM, and vice versa. It appears that there are more absolute biases in LR than NARRM over both western and eastern U.S. Using the longitude 100°W as the divide, the average absolute bias differences (positive indicates NARRM has less overall absolute bias than LR) are 22.8 mm/yr and 0.9 mm/yr over the western and eastern U.S., respectively,

for annual mean runoff, and are 21.6 mm/yr and 18.5 mm/yr over the western and eastern U.S., respectively, for annual mean
600 evapotranspiration. Comparing to LR, NARRM can thus help reduce simulation biases in hydrologic variables.
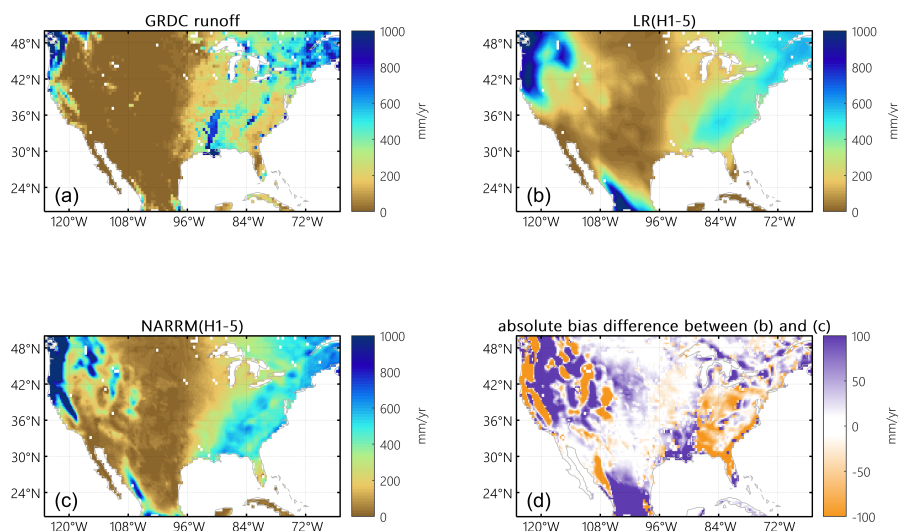
Figure 27. Annual mean runoff from (a) GRDC, (b) LR, (c) NARRM simulation, as well as (d) the differences of absolute biases between LR
and NARRM (former subtracting latter), a positive value suggests the absolute bias in the LR simulation is larger than that in the NARRM,
and vice versa.

5.2.3 Streamflow

Streamflow simulations are typically affected by multiple sources of uncertainties, such as the biases in the simulated runoff, the
uncertainties in the river model parameters (e.g., river network topology, channel geometry, Manning's roughness coefficients),
and water demand data (Li et al., 2013, 2015a, b; Zhou et al., 2020). For river network topology, the half-degree-resolution and
605 1/8th-degree-resolution river network data are used for the LR and NARRM simulations, respectively, as shown in Figs. 29ab.
It is expected that a higher resolution river network data can represent rivers more smoothly, hence more realistically. Another
benefit of higher resolution river network data is to enable more extensive streamflow validation. Terrestrial water fluxes, par-
ticularly surface runoff and streamflow, are dominated by gravity and controlled by topography, hence mostly follow irregular
watershed boundaries. In most land surface and ESMs, including E3SM, regular lat/lon grids are used to resolve spatial hetero-
610 geneity for both runoff and river processes to be compatible with the other land and atmospheric components (Lawrence et al.,
2019; Golaz et al., 2022). Both the magnitude and timing of streamflow at each river gauge are dominated by the corresponding
upstream drainage area. Streamflow simulations are thus largely affected by the discrepancies between the watershed bound-
aries and regular lat/lon grids. These discrepancies can be significantly reduced with higher resolution river network data. For
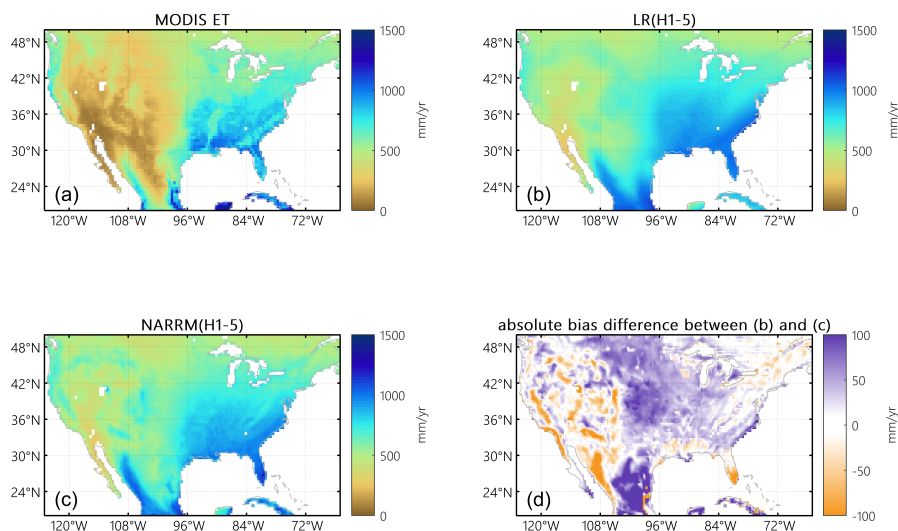
Figure 28. Annual mean evapotranspiration from (a) MODIS, (b) LR, (c) NARRM simulation, as well as (d) the differences of absolute biases between LR and NARRM (former subtracting latter), a positive value suggests the absolute bias in the LR simulation is larger than that in the NARRM, and vice versa.

example, in this study a 10% discrepancy threshold is used to select the river gauges for validating streamflow simulations, i.e.,
615 the relative difference between the real upstream drainage area of a river gauge and that estimated from a lat/lon grid-based river network should not exceed 10%. Over the NA domain, 615 river gauges satisfy the requirement for the half-degree-resolution river network, whilst 2924 river gauges satisfy the requirement for the 1/8th-degree-resolution river network. There are 563 river gauges simultaneously satisfy the requirement for both resolutions.

Figure 29c shows the comparison between the annual mean observed and simulated streamflow over these 563 river gauges.
620 Overall, both simulations produce the long-term average streamflow reasonably well across these gauges. NARRM performs noticeably better (closer to the red 1:1 line) for the top four gauges with largest discharges. An additional analysis (figure not shown) indicates that LR produces greater absolute bias than NARRM in 330 out of 563 gauges (about 60%) in the streamflow simulation. Interestingly, it appears that the over- and underestimation of JJA and DJF streamflow are concentrated in the western and eastern U.S., respectively, for both LR and NARRM, as shown in Figs. 30a–d. Figures 30ef display the difference
625 in absolute biases between LR and NARRM (former subtracting latter) at individual river gauges for the JJA and DJF seasons. Positive difference (indicating greater bias in LR than in NARRM, purple color) dominates over most gauges in eastern US during JJA and over the CONUS during DJF. Taken together, Figs. 29 and 30 suggest an overall better performance of NARRM, despite all the uncertainties.
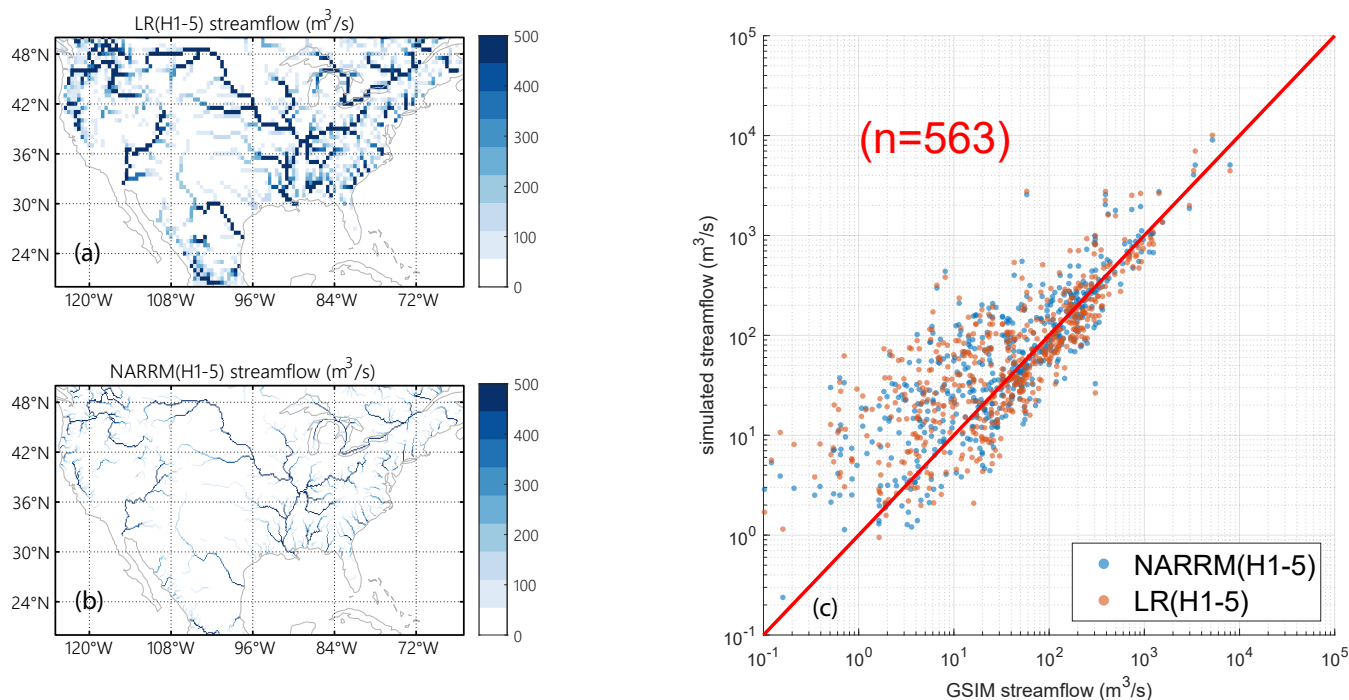
Figure 29. Simulated annual mean streamflow at 563 river gagues in NA compared against The Global Streamflow Indices and Metadata (GSIM) database. (a) River network used in LR, demonstrated by the mean annual discharge; (b) river network used in NARRM, demonstrated by the mean annual discharge; (c) simulated annual mean streamflow against observed for LR and NARRM.

5.3 Land-atmosphere coupling

630   Accurate representation of the interactive processes between the land surface, planetary boundary layer (PBL), and clouds and precipitation is an ongoing challenge for current state-of-art climate models. In this Section, we assess the land-atmosphere (L-A) coupling in LR (H1-5), LR (A1-3), NARRM (H1-5), and NARRM (A1-3) using the 9-yr warm-season observations at the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site, following Tao et al. (2021). Before the detailed analysis on L-A coupling, we first examine the seasonal variations of daytime mean surface heat fluxes from May to August

635   during 2004-2012. As shown in Fig. 31, the evaporative fraction (EF) is in general underestimated in model simulations except for LR (H1-5). The much lower simulated EF compared with the ARM observations is mainly attributed to a large negative bias in surface latent heat fluxes (LH). Different from the other simulations, the surface sensible heat fluxes (SH) is significantly underestimated in LR (H1-5) from May to early July. As a result, the simulated daytime mean EF is higher than that from the observations. Overall, the surface state and fluxes are relatively better reproduced in the historical runs than in the AMIP

640   runs, where both LR (A1-3) and NARRM (A1-3) show a significant negative bias in LH and EF persisting since July. This is surprising and requires further analyses, which is beyond the scope of this paper. In the following, we focus on two local convective regimes and diagnose model behaviors using the Local Coupling metrics (Santanello et al., 2018).
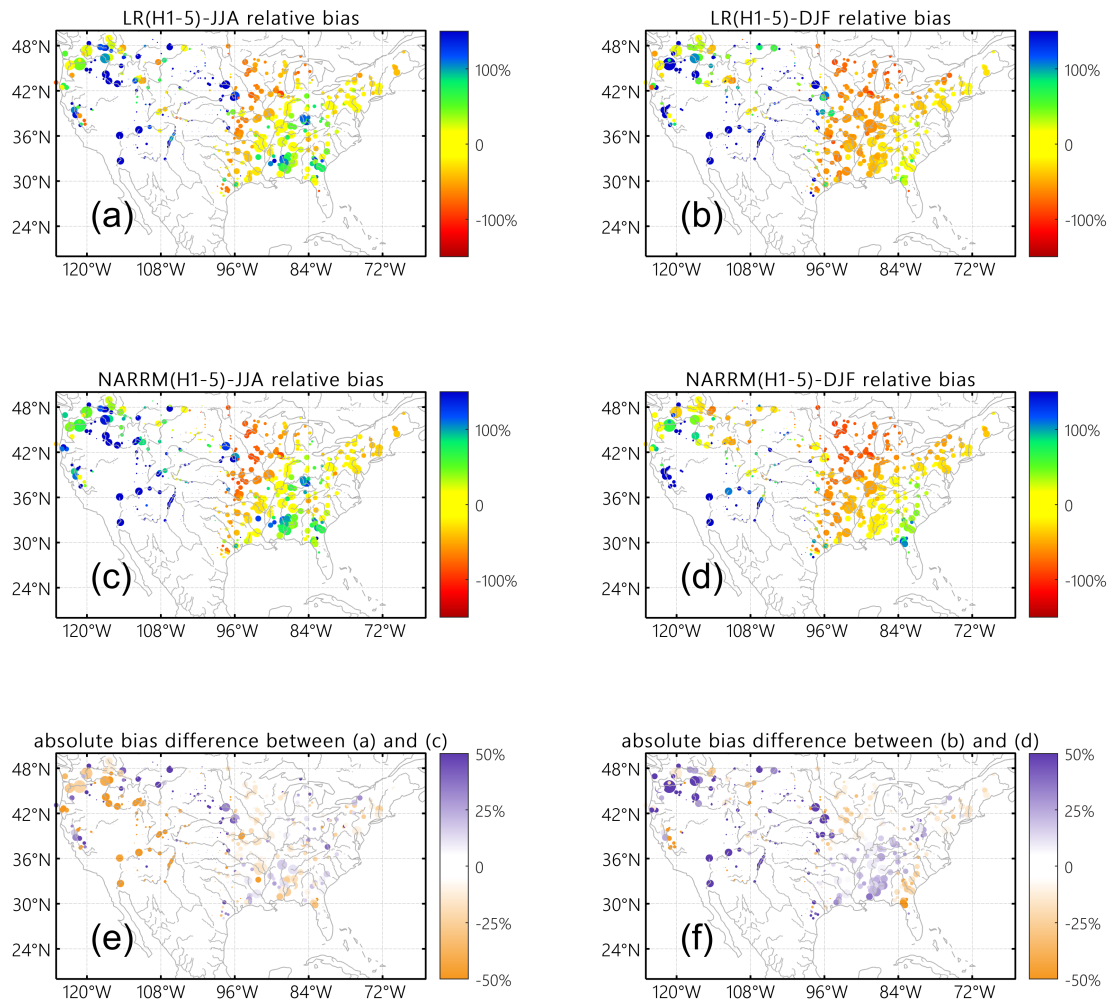
Figure 30. Simulated JJA (left panel) and DJF (right panel) mean streamflow at CONUS river gagues compared against The Global Streamflow Indices and Metadata (GSIM) database. (a) and (b), relative bias of LR; (c) and (d), relative bias of NARRM; (e) the difference of absolute relative bias between LR (absolute value of a) and NARRM (absolute value of c) for the JJA season, positive value (purple) indicates LR has greater absolute bias than NARRM; (f) same as e but for the DJF season

During the selected 9-year period, 165 and 154 clear-sky days are classified from LR (A1-3) and NARRM (A1-3), respectively (Table A1). This is double the 66 clear-sky days identified from ARM observations. However, the occurrence frequency of shallow cumulus (ShCu) days is much lower in these AMIP runs compared with that observed. Only 6 and 5 ShCu days are identified in LR (A1-5) and NARRM (A1-5), respectively (not shown). For the historical runs, the number of selected clear-sky days from both LR (H1-5) and NARRM (H1-5) are comparable to that observed but the occurrence frequency of ShCu days is still low. As we are targeting for a statistical and climatological comparison between the long-term ARM data and climate model simulations, we extend the analysis period to 1980-2012 for model simulations on ShCu days due to the limited sample size between 2004 and 2012.

Figure 32 shows the composite clear-sky day mixing diagrams (Santanello et al., 2009), which relates the conservative variables, potential temperature (θ) and total water specific humidity (q), to the water and energy budgets and the growth of planetary boundary layer (PBL). The coevolution of Lvq and Cpθ (0730 to 1730 LST) is decomposed by vector components that represent the integrated fluxes of heat and moisture from the land surface (V_{sfc}), the advection (V_{adv}) and the entrainment at the PBL top (V_{ent} as a residual). Six metrics are derived from these diagrams and summarized in Table 5. In general, although the differences among various model simulations are minor, several common model biases are noted when compared with the ARM observations. For example, the model simulated clear-sky days are featured with too warm and too dry conditions in the early morning (0730 LST). The θ_{sfc} in the two AMIP runs and historical runs, both LR and NARRM, are about three and two times of that observed, respectively. The high θ_{sfc} indicates that more energy at the surface goes to heating rather than moistening. Moreover, the E_{LH} is significantly overestimated in the model simulations, which is about five (three) times of that observed in the two AMIP runs (historical runs). The much higher simulated E_{LH} suggests that the entrainment heating and drying dominate the surface fluxes on the simulated clear-sky days, which supports rapid and deep PBL growth in models. Different from the observations, the simulated advection tends to cool and dry the mixed layer, but the overall impact is much smaller compared with those from the surface and entrainment.

Figure 33 shows the daytime evolution composites of PBL, lifting condensation level (LCL), and LCL deficit (PBL top height minus LCL) on clear-sky and ShCu days. Both PBL and LCL on model simulated clear-sky days are much higher than that on the ARM observed clear-sky days. The model behaviors are in general consistent among different simulations, except that the bias of LCL is significantly lower in LR (H1-5) compared with the others. In models, the PBL grows rapidly after the sunrise on clear-sky days, corresponding to the large warm and dry air entrainment that dominate the PBL budget (Fig. 32). But the too warm and too dry early-morning surface conditions lead to an even higher LCL on model simulated clear-sky days. The PBL never reaches the LCL, with a negative LCL deficit throughout the day, which supports clear skies. The diurnal evolution of LCL on the ARM-observed ShCu days is similar to that on clear-sky days but the development of PBL is much more vigorous. As a result, the PBL is deep enough to touch the LCL for cloud formation around noon. Different from that observed, the daytime evolution of PBL is much weaker on ShCu days than that on clear-sky days in all model simulations. However, the decrease in LCL from clear-sky days to ShCu days is even greater, where the growth of PBL is high enough to touch the LCL for cloud formation. Note that the models simulate a positive LCL deficit at around 0900 LST, a few hours earlier than that in the observations. To summarize, ShCu forms as a result of strong surface SH fluxes that drives the rapid

development of PBL in observations while in models, ShCu results from a relatively more humid lower troposphere that leads to a lowered LCL. Differences among various model simulations are pretty minor.
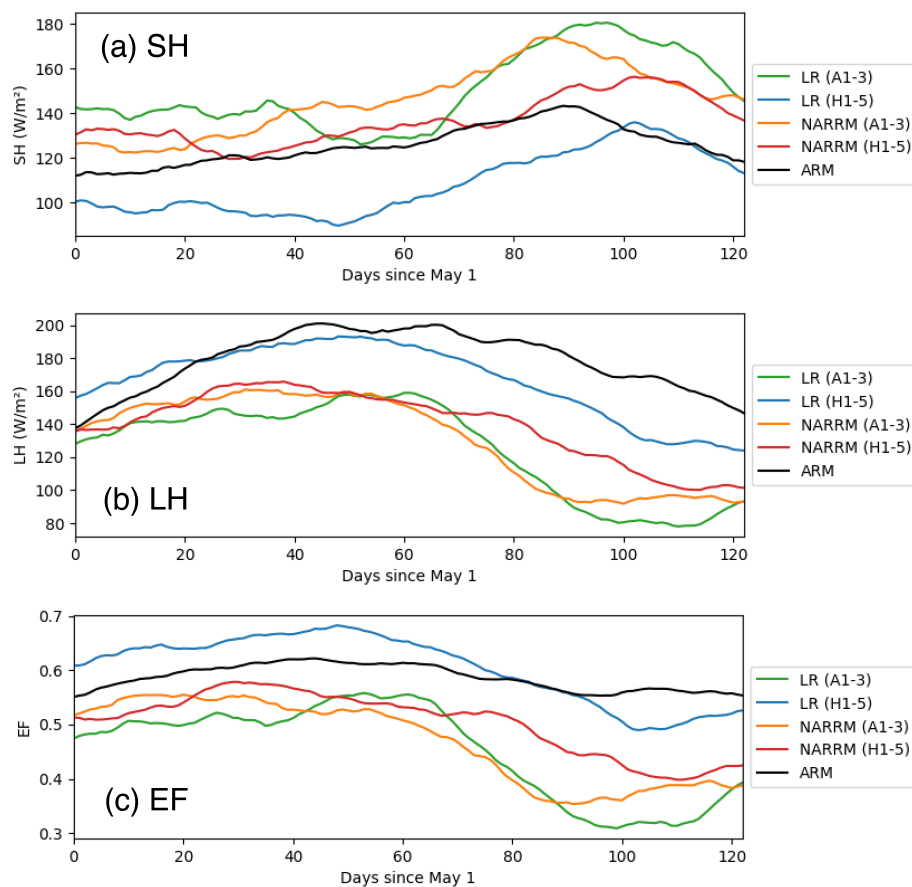
Figure 31. The seasonal variation of 2004-2012 daytime mean (0600-1800 LST): (a) surface sensible heat flux (SH), (b) surface latent heat flux (LH), and (c) surface evaporative fraction (EF, defined as [LH/(LH+SH)]) from ARM observations (black), LR (A1-3) (green), LR (H1-5) (blue), NARRM (A1-3) (orange), and NARRM (H1-5). A moving average of 30 days is applied to smooth out short-term fluctuations and highlight longer-term trends.

6   Conclusions and discussions

A primary Earth system model (ESM) advancement is to represent the spatially continuous world more realistically on discretized grids, which often requires constantly increasing the finest scale of explicitly resolved processes within the computational limit. Before uniformly high-resolution global models solve their severe computational challenge for climate simulation
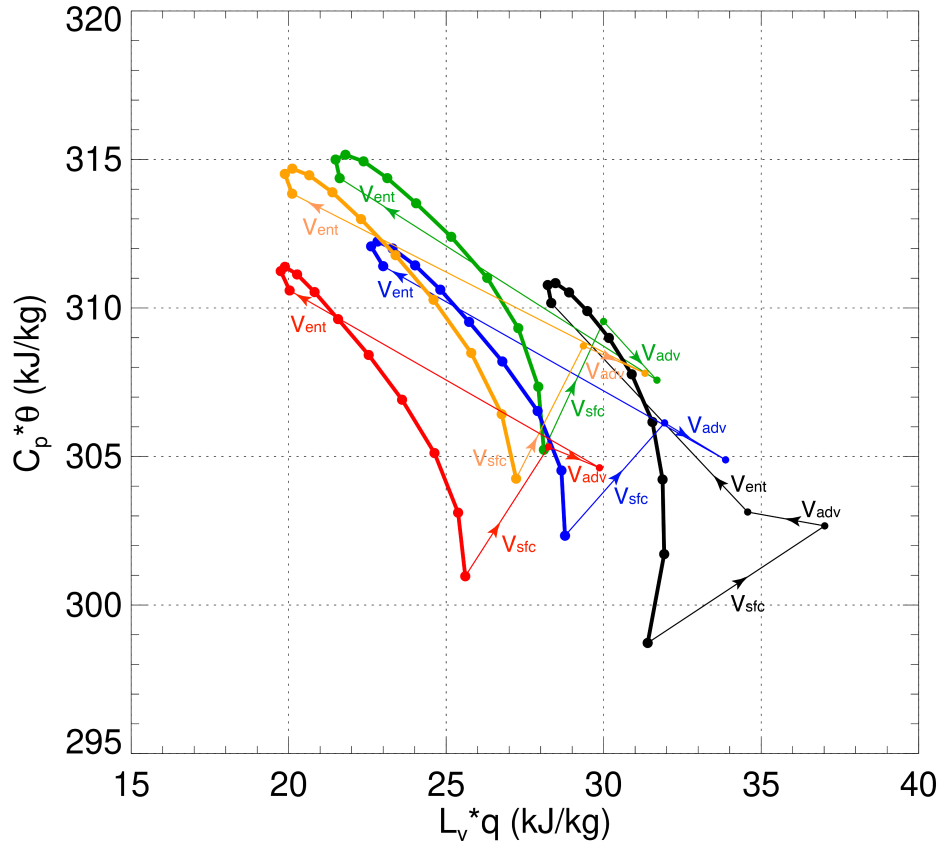
Figure 32. Clear-sky day mixing diagram of the PBL conservative variables, Lvq versus Cpθ, during the daytime evolution from ARM observations (black), LR (A1-3) (green), LR (H1-5) (blue), NARRM (A1-3) (orange), and NARRM (H1-5) (red). Dots denote the composite hourly means from 0730 to 1730 local time. The text annotations depict the vector component contributions from surface (Vsfc), advection (Vadv) and entrainment fluxes (Vent) to the evolution.

Table 5. The surface (β_{sfc}) and entrainment (β_{ent}) Bowen ratios, the entrainment ratio of heat (E_{SH}) and moisture (E_{LH}), and the advective flux ratio of heat (A_{SH}) and moisture (A_{LH}) from the ARM observations, LR (A1-3), LR (H1-5), NARRM (A1-3), and NARRM (H1-5) on clear-sky days. The flux values (W/m^2) are derived using the mixing diagram theory and surface, advection, and entrainment flux vectors depicted in Figure 32.

Metrics	Obs.	LR (A1-3)	LR (H1-5)	NARRM (A1-3)	NARRM (H1-5)
$\beta_{sfc} = \mathrm{SH}_{sfc}/\mathrm{LH}_{sfc}$	0.70	2.28	1.20	2.09	1.65
$\beta_{ent} = \mathrm{SH}_{ent}/\mathrm{LH}_{ent}$	-1.13	-0.68	-0.60	-0.54	-0.61
$E_{SH} = \mathrm{SH}_{ent}/\mathrm{SH}_{sfc}$	1.79	1.57	1.72	1.35	1.37
$E_{LH} = \mathrm{LH}_{ent}/\mathrm{LH}_{sfc}$	-1.11	-5.31	-3.44	-5.24	-3.72
$A_{SH} = \mathrm{SH}_{adv}/(\mathrm{SH}_{sfc}+\mathrm{SH}_{ent})$	0.04	-0.18	-0.12	-0.09	-0.07
$A_{LH} = \mathrm{LH}_{adv}/(\mathrm{LH}_{sfc}+\mathrm{LH}_{ent})$	3.97	-0.21	-0.25	-0.22	-0.22
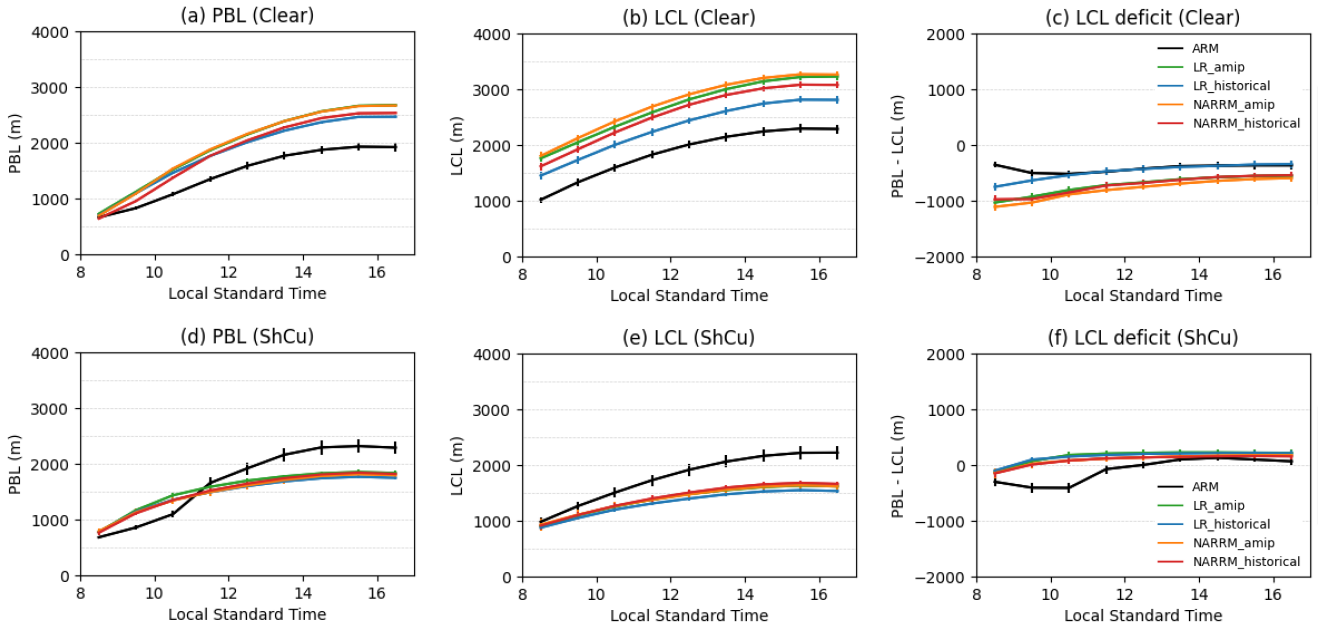
Figure 33. Composite daytime evolution of (a) PBL, (b) LCL, and (c) LCL deficit (PBL minus LCL) from ARM observations (black), LR (A1-3) (green), LR (H1-5) (blue), NARRM (A1-3) (orange), and NARRM (H1-5) (red) on clear-sky days. (d–f) are as (a–c) but on shallow cumulus days.

campaigns, the multiresolution ESM (e.g., regionally refined model (RRM)) is a natural alternative for these campaigns. Nev-ertheless, it has been over a decade since such a multiresolution method (e.g., Ringler et al., 2008) was proposed.

To our knowledge, this is the first study with a global ESM that has accomplished the CMIP6 climate simulation campaign with a fully coupled RRM configuration — a potentially significant step in the long journey of improving the explicitly resolved resolution of climate simulations. The key to this success is the application of the hybrid timestep strategy (i.e., merging high-resolution dynamics timestep with the low-resolution physics timestep) in the atmosphere model, which mitigates the negative impacts caused by the persistent poor scale-aware problem of atmospheric physics in a multi-scale framework (e.g., RRM). The powerful aspect of RRM is that it typically only costs ∼10–20% of the globally uniform high-resolution model, substantially reducing the computational burden of production simulations. This is particularly important for high-resolution ensemble simulations, which are necessary to account for the internal variability of the climate system, but whose cost would otherwise be prohibitive. On the global scale, we show that NARRM well reproduces the LR climate. Within the high-resolution domain (i.e., North America), NARRM displays more improvements than deteriorations relative to LR. Furthermore, some of the NARRM improvements (e.g., marine shallow cumulus clouds at California and mixed-phase clouds near Arctic) are attributable to the better captured coupling processes, highlighting the strength of refining multiple components over a single component. The main detailed findings are as follows.

– The new dry baroclinic idealized test (Hughes and Jablonowski, 2023) allows us to test the NARRM grid with the stand-alone atmospheric dynamical core and confirms that the NARRM mesh is numerically stable, and the results are reasonable compared to the LR and HR grids (Fig. 4).

– By employing the EAM hybrid timestep method, NARRM, without retuning physics parameters, successfully matches the global climate (including climatology, time series, and climate sensitivity and feedback) simulated by LR (Figs. 7, 9, 10, 22).

– Within the high-resolution region over the CONUS, precipitation and clouds are largely improved in NARRM compared to LR (Figs. 11ac, 12, 13) due to the better topography in NARRM (Fig. 8) and/or reduced sea surface temperature biases.

– Refining the atmospheric grid spacing from 100 km to 25 km is not adequate to improve the diurnal propagation of organized MCSs at the CONUS (Fig. 15).

– NARRM retains the LR performance of aerosol simulations on the global scale and regional mean basis, without re-tuning of the scale-dependent aerosol emissions. Over the refined mesh, NARRM improves the simulated aerosol spatial variability and predictions of extreme polluted cases (e.g., the upper tail of AOD distribution) (Fig. 18). On the other hand, the refined grid resolution does not eliminate the high biases in aerosol loadings and effective radiative forcing, inherited from the LR model.

- NARRM generally simulates better cloud covers than LR for both liquid and ice phase clouds. Over land (e.g., western NA and Greenland), this improvement is likely related to topography, whereas over ocean, it is attributed to air-sea interactions.

- NARRM produces a comparable global-mean cloud feedback to LR but a less positive cloud feedback over the NA (Fig. 22). The reduction in cloud feedback there mainly relates to the shortwave component. The total cloud feedbacks over coastal California does not change much due to the compensation between shortwave and longwave components (Fig. 23).

- While both the LR and NARRM simulations are able to capture to a large extent the spatial and statistical distributions of the observed extratropical cyclone (ETC) activities, NARRM shows a particularly better skill in simulating the ETC activities along the oceanic storm tracks and over the mountain range to the east of the Rocky Mountains (Fig. 24). NARRM in coupled mode outperforms all other configurations (LR and uncoupled NARRM) in simulating the shape and orientation of the oceanic storm tracks within the NARRM high-resolution domain, due to the coupling with the refined ocean surrounding North America. NARRM in general produces more ETCs than LR, and overestimates the total number of cyclones compared to the ERA5 reanalysis. But probably more importantly, for intense cyclones or rapidly developing ones, the NARRM simulations are in close agreement with the observations while the LR simulations mismatch by a significant margin (Fig. 25).

- NARRM appears to better represent the spatial variability in land hydrologic processes by resolving the land features more realistically over the western US (Figs. 27, 28). With higher grid resolution, NARRM can better capture surface topography which dominates surface water flows across hillslopes and through rivers, hence not only improving the river model performance but also providing more precise river gauge geo-referencing information for streamflow validations (Figs. 29, 30).

- NARRM provides enhanced winter (DJF) climatological representation of the spatial variability of snow water equivalent (SWE) across the CONUS relative to LR (Fig. 26ab) as a result of higher SWE magnitudes and more granular spatial structures in NARRM. Certain biases (e.g., peak water volume) of snowpack are reduced in NARRM compared with LR (Fig. 26c).

- Over the ARM SGP site during warm seasons, the surface conditions are warm and dry on the model simulated clear-sky days with overestimation in both the PBL height and the LCL (Figs. 32, 33), while the ShCu days in models result from a much moister environment compared with that in the observations (Fig. 33). In general, the surface properties and fluxes are relatively better reproduced in the historical runs than in the AMIP runs (Fig. 31), with limited impact of resolutions.

Besides the NARRM configuration illustrated in the present study, E3SMv2 has been successfully run with RRM meshes with finer grids located at other regions (Antarctic, Arctic, and Southeast Pacific). We expect that the hybrid timestep strategy is a general approach that can be applied to these RRMs to simulate high-resolution climate at different areas. With that in mind,

we streamlined the process of creating new RRM configurations to facilitate broader RRM applications in the next phase of the E3SM project. Depending on the goal of RRM simulations, further improvements over the refined domain can be achieved via additional parameter tuning. In that case, nudging the outside coarser domain may be necessary to avoid severe degradations to the climate there. Such nudging capability is available in E3SM (e.g., Tang et al., 2019) and one has the option to nudge towards the data from reanalysis product or low-resolution E3SM simulation. Lastly, we highlight that this paper serves as an overview of the NARRM atmosphere, land, and river models. More in-depth analysis is planned to be reported in follow-up papers.

Code and data availability. The E3SM code used in this work is available at https://doi.org/10.5281/zenodo.7343230 and on GitHub at https://github.com/E3SM-Project/E3SM, including a maintenance branch (`maint-2.0`; https://github.com/E3SM-Project/E3SM/tree/maint-2.0) which has been created to reproduce these simulations.

Complete native model output as well as nudging simulations' climatology data are accessible directly at the National Energy Research Scientific Computing Center (NERSC): https://portal.nersc.gov/archive/home/projects/e3sm/www/WaterCycle/E3SMv2/LR and https://portal.nersc.gov/archive/home/projects/e3sm/www/WaterCycle/E3SMv2/NARRM for low-resolution and NARRM simulations, which are documented at https://e3sm-project.github.io/e3sm_data_docs. A subset of the native output is also available through the DOE Earth System Grid Federation (ESGF) at https://esgf-node.llnl.gov/search/e3sm/?model_version=2_0. Data reformatted following CMIP conventions will also be available through ESGF at https://esgf-node.llnl.gov/projects/e3sm.

Performance data and scripts are located at https://github.com/E3SM-Project/perf-data/tree/archive/v2-narrm-perf-study/v2-narrm.

Appendix A
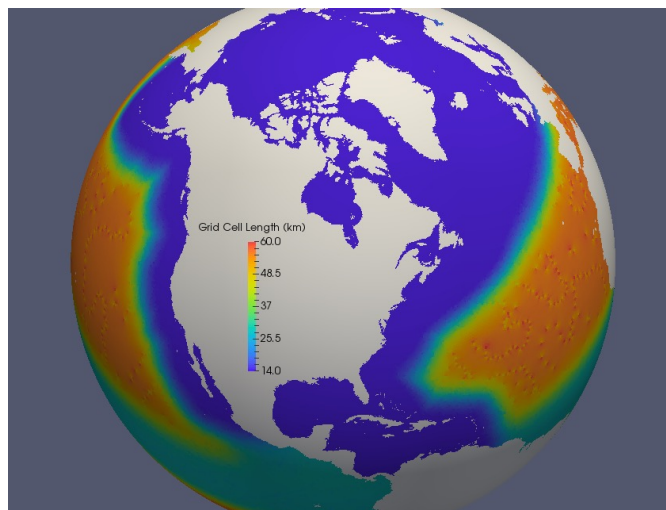
Figure A1. North American RRM (NARRM) grids for ocean and sea ice.
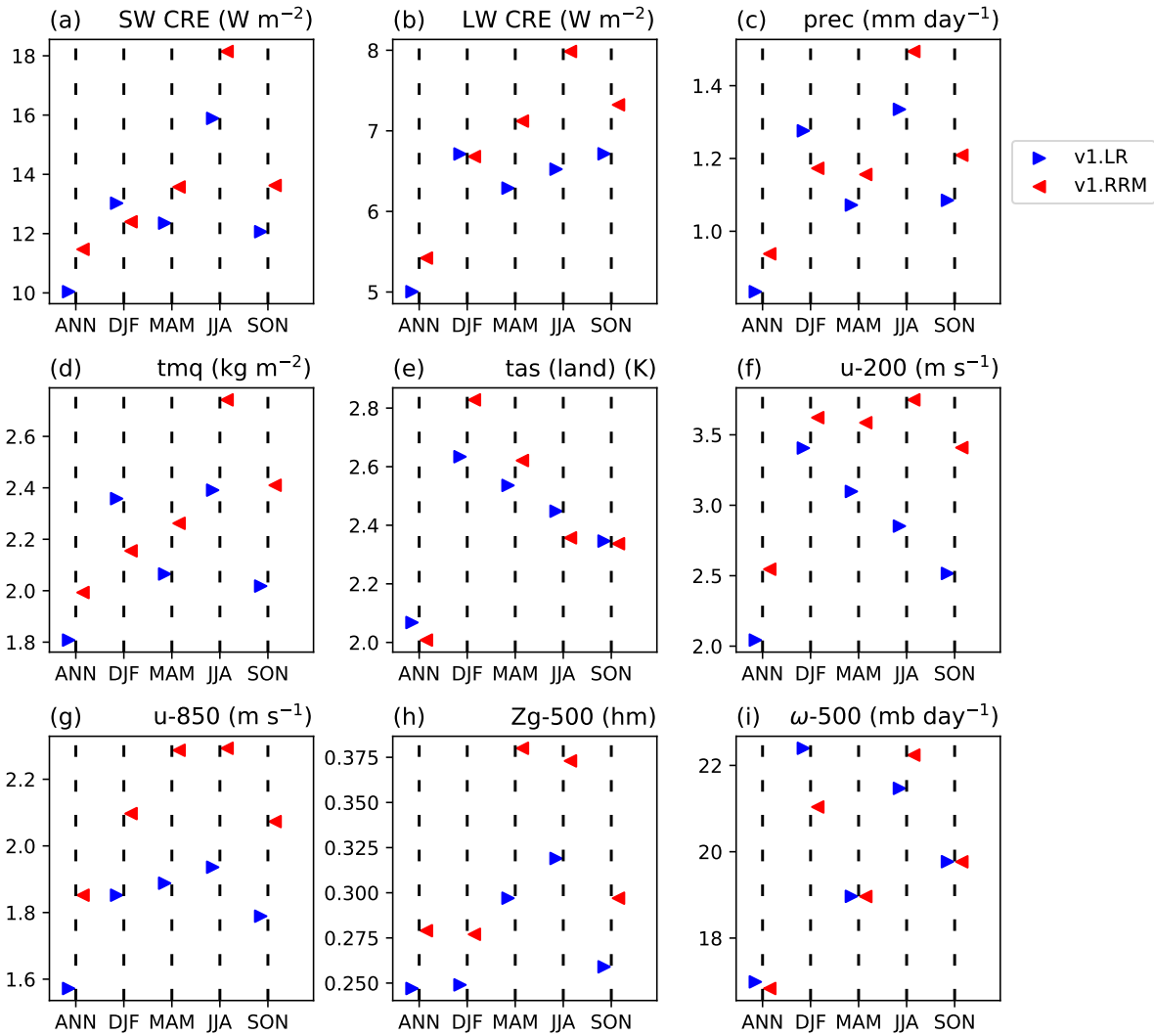
Figure A2. Same as Fig. 11, but for the global spatial RMSE of model climatology from the EAMv1 LR (blue triangles) and the EAMv1 RRM (red triangles) with high-resolution grids over the CONUS and both physics parameters and timesteps tuned for high-resolution. The details of these two EAMv1 simulations are documented in Tang et al. (2019).
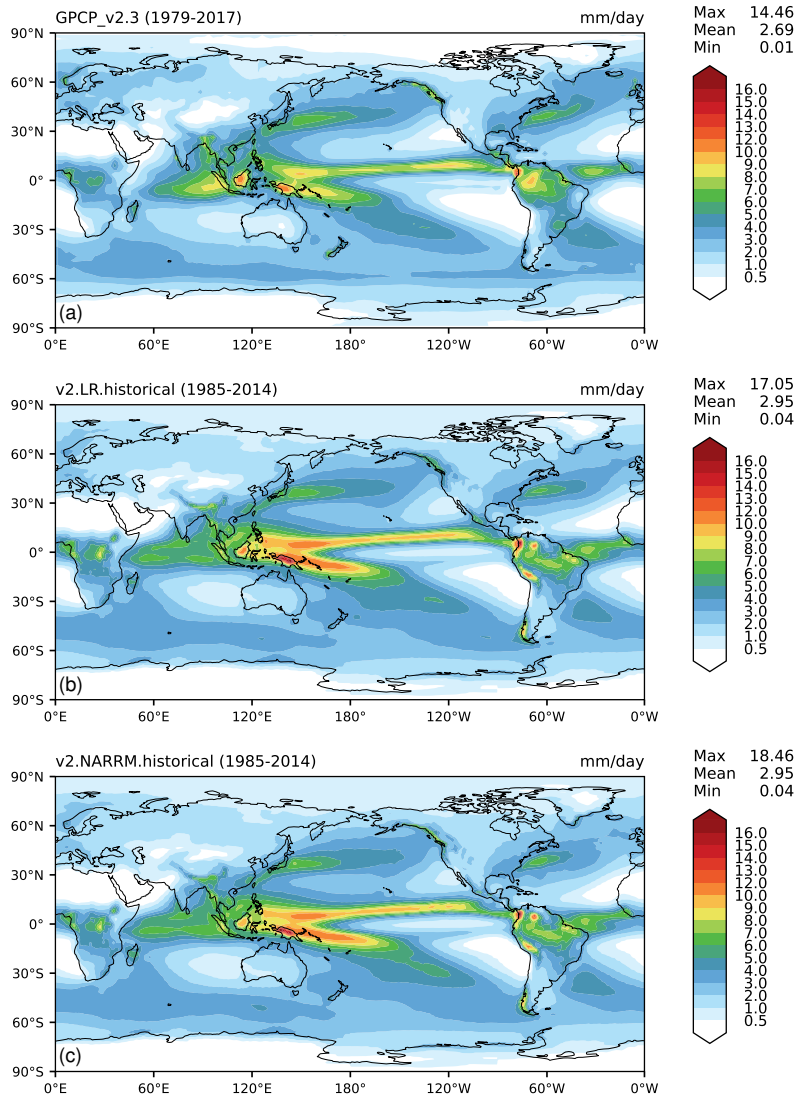
Figure A3. Comparison of global annual mean precipitation geographic patterns from (a) GPCP2.3, (b) LR, and (c) NARRM historical ensemble means
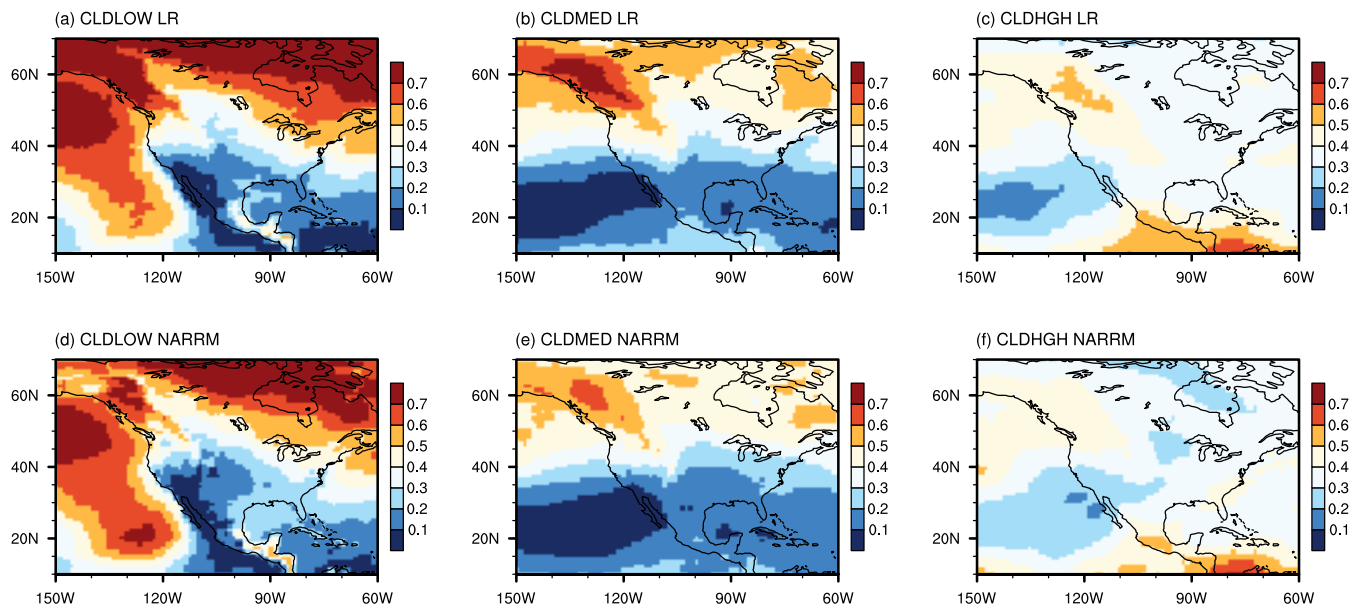
Figure A4. Cloud fractions over North America simulated by the nudged LR (top) and NARRM (bottom) for low clouds (CLDLOW; left), middle clouds (CLDMED; middel), and high clouds (CLDHGH; right).
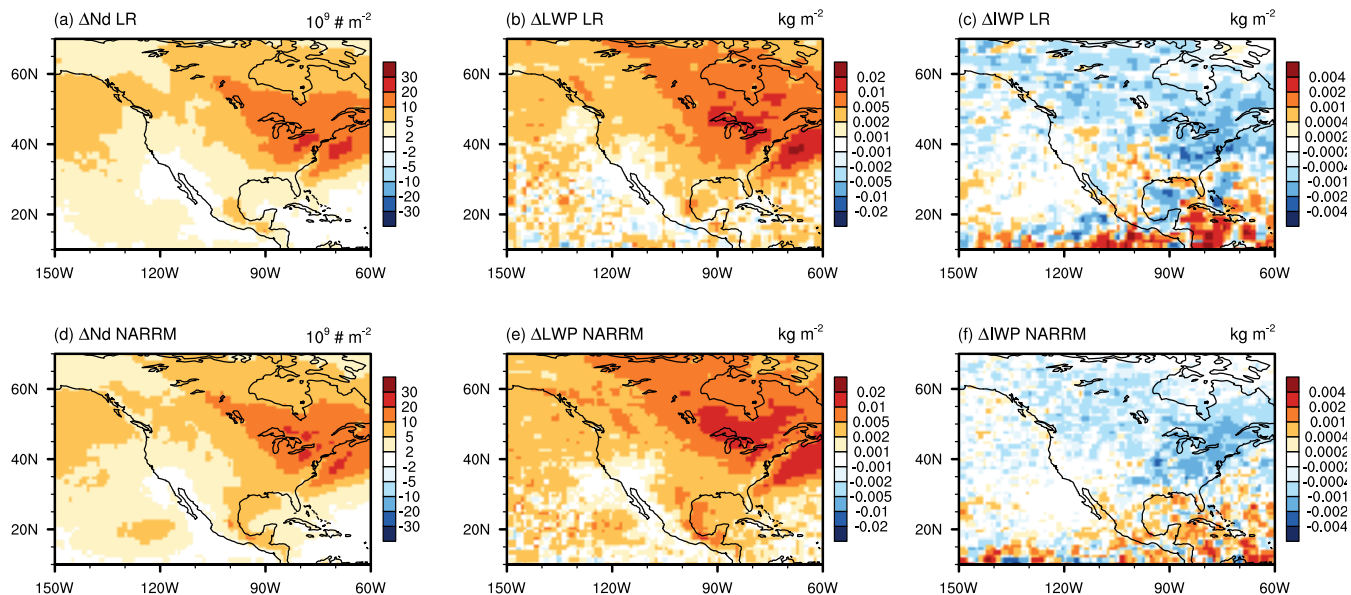
Figure A5. PD-PI (present-day (PD) minus pre-industrial (PI)) changes in cloud droplet number concentration (left), liquid water path (LWP), and ice water path (IWP) at North America calculated from the nudged LR (top) and NARRM (bottom) simulations.

Table A1. Definition criteria and sample size of the clear-sky regime (Clear) and shallow cumulus regime (ShCu) based on the ARM observations and four different model simulations.

Regime	Definition criteria	Obs.	LR (A1-3)	LR (H1-5)	NARRM (A1-3)	NARRM (H1-5)
Clear	**Obs: Analysis period (2004-2012)** • Precipitation rate = 0 mm/hr at all 24 hours • Between 0800 and 1600 LST, total cloud fraction ≤ 15%, low-level and mid-level cloud fraction ≤ 5%, and high-level cloud fraction ≤ 10% **Model: Analysis period (2004-2012)** • Precipitation rate < 0.1 mm/hr at all 24 hours • Between 0800 and 1600 LST, total cloud fraction ≤ 15%, low-level and mid-level cloud fraction ≤ 5%, and high-level cloud fraction ≤ 10%	66	165	86	154	66
ShCu	**Obs: Analysis period (2004-2012)** • Precipitation rate = 0 mm/hr at all 24 hours • Cloud tops < 4 km and cloud bases gradually rise with time over the day • Above 4 km, there is usually no cloud or cloud fraction < 5%, except on a few days when there is some high cirrus above 10 km • Satellite images of ShCu days identified based on Active Remote Sensing of Clouds data and the Total Sky Imager are examined manually to ensure that the cloud field develops homogeneously and is not affected by other large-scale weather phenomena **Model: Analysis period (1980-2012)** • Precipitation rate < 0.25 mm/hr at all 24 hours • Diurnal maximum hourly low-level cloud fraction between 5% and 70%, and between 1000 and 1800 LST • Between 0000 and 0600 LST, low-level cloud fraction < 5% • Diurnal maximum hourly mid-level cloud fraction ≤ 10% and is lower than that of low-level cloud fraction at all 24 hours	48	34	66	23	48

Author contributions. As part of the E3SM water cycle group co-led by JCG and LPVR, QT led the Regionally Refined Model (RRM) development team with significant contributions from JCG, MAT, WL, BRH, PAU, and AMB, and help from OG, JDW, CZ, CSZ, ELR, AFR, MEM, RLJ, AVDV, PMC, and GB. The RRM configuration leverages the low-resolution model development effort from TZ, KZ, and XZ. QT and JCG designed the paper scope. QT, JCG, MAT, AMB, KZ, BS, and RMF carried out the model simulations with assistance from AM, NDK, and RLJ. QT led the analysis and the manuscript writing with significant contributions from JCG, MAT, WL, AMB, TZ, KZ, YZ, MZ, MW, HW, CT, AMR, YQ, HYL, and YF. All co-authors contributed to the manuscript.

Competing interests. At least one of the (co-)authors is a member of the editorial board of "Geoscientific Model Development".

Acknowledgements. This research was supported as part of the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research (BER). E3SM production simulations were performed on a high-performance computing cluster provided by the BER Earth System Modeling program and operated by the Laboratory Computing Resource Center at Argonne National Laboratory, and the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Developmental simulations were performed using BER Earth System Modeling program's Compy computing cluster located at Pacific Northwest National Laboratory and NERSC machine cori-knl. Additional developmental simulations, as well as post-processing and data archiving of production simulations used resources of NERSC.

Lawrence Livermore National Laboratory (LLNL) is operated by Lawrence Livermore National Security, LLC, for the U.S. DOE, National Nuclear Security Administration under Contract DE-AC52-07NA27344. Support has been received from the LLNL LDRD projects 22-ERD-008, "Multiscale Wildfire Simulation Framework and Remote Sensing", the DOE Atmospheric Radiation Measurement (ARM) and the DOE Atmospheric System Research (ASR) program. Data from the U.S. DOE as part of the ARM Climate Research Facility Southern Great Plains site were used. Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This work was supported by the U.S. DOE through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001).

References

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D.-B.: The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation, Atmosphere, 9, 138, https://doi.org/10.3390/atmos9040138, 2018.

Bambach, N. E., Rhoades, A. M., Hatchett, B. J., Jones, A. D., Ullrich, P. A., and Zarzycki, C. M.: Projecting climate change in South America using variable-resolution Community Earth System Model: An application to Chile, International Journal of Climatology, pp. 1–29, https://doi.org/10.1002/joc.7379, 2021.

Bengtsson, Y., Hodges, K. I., and Roeckner, R.: Storm tracks and climate change, Journal of Climate, 19, 3518–3543, https://doi.org/10.1175/JCLI3815.1, 2006.

Beres, J. H., Alexander, M. J., and Holton, J. R.: A Method of Specifying the Gravity Wave Spectrum above Convection Based on Latent Heating Properties and Background Wind, Journal of the Atmospheric Sciences, 61, 324 – 337, https://doi.org/10.1175/1520-0469(2004)061<0324:AMOSTG>2.0.CO;2, 2004.

Blender, R. and Schubert, M.: Cyclone tracking in different spatial and temporal resolutions, Monthly Weather Review, 128, 377–384, https://doi.org/10.1175/1520-0493(2000)128%3C0377:CTIDSA%3E2.0.CO;2, 2000.

Bogenschutz, P. A., Lee, H.-H., Tang, Q., and Yamaguchi, T.: Combining regional mesh refinement with vertically enhanced physics to target marine stratocumulus biases as demonstrated in the Energy Exascale Earth System Model version 1, Geoscientific Model Development, 16, 335–352, https://doi.org/10.5194/gmd-16-335-2023, 2023.

Bradley, A. M., Bosler, P. A., and Guba, O.: Islet: interpolation semi-Lagrangian element-based transport, Geosci. Model Dev., 15, 6285–6310, https://doi.org/10.5194/gmd-15-6285-2022, 2022.

Burakowski, E. A., Tawfik, A., Ouimette, A., Lepine, L., Zarzycki, C., Novick, K., Ollinger, S., and Bonan, G.: Simulating surface energy fluxes using the variable-resolution Community Earth System Model (VR-CESM), Theoretical and Applied Climatology, 138, 115–133, https://doi.org/10.1007/s00704-019-02785-0, 2019.

Caldwell, P. M., Mametjanov, A., Tang, Q., Van Roekel, L. P., Golaz, J.-C., Lin, W., Bader, D. C., Keen, N. D., Feng, Y., Jacob, R., Maltrud, M. E., Roberts, A. F., Taylor, M. A., Veneziani, M., Wang, H., Wolfe, J. D., Balaguru, K., Cameron-Smith, P., Dong, L., Klein, S. A., Leung, L. R., Li, H.-Y., Li, Q., Liu, X., Neale, R. B., Pinheiro, M., Qian, Y., Ullrich, P. A., Xie, S., Yang, Y., Zhang, Y., Zhang, K., and Zhou, T.: The DOE E3SM Coupled Model Version 1: Description and Results at High Resolution, Journal of Advances in Modeling Earth Systems, 11, 4095–4146, https://doi.org/10.1029/2019MS001870, 2019.

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., Beydoun, H., Taylor, M., Bertagna, L., Bradley, A. M., Clevenger, T. C., Donahue, A. S., Eldred, C., Foucar, J., Golaz, J.-C., Guba, O., Jacob, R., Johnson, J., Krishna, J., Liu, W., Pressel, K., Salinger, A. G., Singh, B., Steyer, A., Ullrich, P., Wu, D., Yuan, X., Shpund, J., Ma, H.-Y., and Zender, C. S.: Convection-Permitting Simulations With the E3SM Global Atmosphere Model, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 544, https://doi.org/10.1029/2021MS002544, e2021MS002544 2021MS002544, 2021.

Cesana, G. and Chepfer, H.: Evaluation of the cloud thermodynamic phase in a climate model using CALIPSO-GOCCP, Journal of Geophysical Research: Atmospheres, 118, 7922–7937, 2013.

Chang, E. K. M.: CMIP5 projection of significant reduction in extratropical cyclone activity over North America, Journal of Climate, 26, 9903–9922, https://doi.org/10.1175/JCLI-D-13-00209.1, 2013.

Chepfer, H., Bony, S., Winker, D., Chiriaco, M., Dufresne, J.-L., and Sèze, G.: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model, Geophysical Research Letters, 35, 2008.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J., Minnis, P., Stubenrauch, C., and Zeng, S.: The GCM-oriented calipso cloud product (CALIPSO-GOCCP), Journal of Geophysical Research: Atmospheres, 115, 2010.

Dang, C., Zender, C. S., and Flanner, M. G.: Intercomparison and improvement of two-stream shortwave radiative transfer schemes in Earth system models for a unified treatment of cryospheric surfaces, The Cryosphere, 13, 2325–2343, https://doi.org/10.5194/tc-13-2325-2019, 2019.

Demory, M., Vidale, P., Roberts, M., Berrisford, P., Strachan, J., Schiemann, R., and Mizielinski, M.: The role of horizontal resolution in simulating drivers of the global hydrological cycle, Climate Dynamics, 42, 2201–2225, https://doi.org/10.1007/s00382-013-1924-4, 2014.

Dennis, J. M., Fournier, A., Spotz, W. F., St-Cyr, A., Taylor, M. A., Thomas, S. J., and Tufo, H.: High-Resolution Mesh Convergence Properties and Parallel Efficiency of a Spectral Element Atmospheric Dynamical Core, The International Journal of High Performance Computing Applications, 19, 225–235, https://doi.org/10.1177/1094342005056108, 2005.

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., St-Cyr, A., Taylor, M. A., and Worley, P. H.: CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, The International Journal of High Performance Computing Applications, 26, 74–89, https://doi.org/10.1177/1094342011428142, 2011.

Devanand, A., Huang, M., Lawrence, D. M., Zarzycki, C. M., Feng, Z., Lawrence, P. J., Qian, Y., and Yang, Z.: Land Use and Land Cover Change Strongly Modulates Land-Atmosphere Coupling and Warm-Season Precipitation Over the Central United States in CESM2-VR, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 925, https://doi.org/10.1029/2019MS001925, e2019MS001925 2019MS001925, 2020.

Di Vittorio, A. V., Chini, L. P., Bond-Lamberty, B., Mao, J., Shi, X., Truesdale, J., Craig, A., Calvin, K., Jones, A., Collins, W. D., Edmonds, J., Hurtt, G. C., Thornton, P., and Thomson, A.: From land use to land cover: restoring the afforestation signal in a coupled integrated assessment–earth system model and the implications for CMIP5 RCP simulations, Biogeosciences, 11, 6435–6450, https://doi.org/10.5194/bg-11-6435-2014, 2014.

Evans, K., Lauritzen, P., Mishra, S., Neale, R., Taylor, M., and Tribbia, J.: AMIP Simulation with the CAM4 Spectral Element Dynamical Core, J. Climate, 26, 689–709, https://doi.org/10.1175/jcli-d-11-00448.1, 2013.

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

Fekete, B., Vorosmarty, C., Hall, F., Collatz, G., Meeson, B., Los, S., Brown De Colstoun, E., and Landis, D.: Islscp ii unh/grdc composite monthly runoff, ORNL DAAC, 2011.

Feng, Y., Wang, H., Rasch, P. J., Zhang, K., Lin, W., Tang, Q., Xie, S., Hamilton, D. S., Mahowald, N., and Yu, H.: Global Dust Cycle and Direct Radiative Effect in E3SM Version 1: Impact of Increasing Model Resolution, Journal of Advances in Modeling Earth Systems, p. 50, https://doi.org/10.1029/2021MS002909, 2022.

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., Fiorino, M., Gleckler, P. J., Hnilo, J. J., Marlais, S. M., Phillips, T. J., Potter, G. L., Santer, B. D., Sperber, K. R., Taylor, K. E., and Williams, D. N.: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I), Bulletin of the American Meteorological Society, 80, 29 – 56, https://doi.org/10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2, 1999.

Geng, Q. and Sugi, M.: Variability of the North Atlantic cyclone activity in winter analyzed from NCEP-NCAR reanalysis data, Journal of Climate, 14, 3863–3783, https://doi.org/10.1175/1520-0442(2001)014%3C3863:VOTNAC%3E2.0.CO;2, 2001.

Gettelman, A., Callaghan, P., Larson, V. E., Zarzycki, C. M., Bacmeister, J. T., Lauritzen, P. H., Bogenschutz, P. A., and Neale, R. B.: Regional Climate Simulations With the Community Earth System Model, Journal of Advances in Modeling Earth Systems, 10, 1245–1265, https://doi.org/10.1002/2017MS001227, 2018.

GISTEMP Team: GISS surface temperature analysis (GISTEMP), NASA Goddard Institute for Space Studies, https://data.giss.nasa.gov/gistemp, dataset accessed 2018-05-24, 2018.

Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-based model for boundary layer clouds. Part I: Method and model description, J. Atmos. Sci., 59, 3540–3551, https://doi.org/10.1175/1520-0469(2002)059<3540:apbmfb>2.0.co;2, 2002.

Golaz, J.-C., Van Roekel, L. P., Zheng, X., and et al.: The DOE E3SM Model Version 2: Overview of the physical model, Journal of Advances in Modeling Earth Systems, in review, 2022.

Greeves, C. Z., Pope, V. D., Stratton, R. A., and Martin, G. M.: Representation of Northern Hemisphere winter storm tracks in climate models, Climate Dynamics, 28, 683–702, https://doi.org/10.1007/s00382-006-0205-x, 2007.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, https://doi.org/10.1029/2003gl018747, 2004.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., Dunne, J. P., Goddard, P., Morrison, A. K., Rosati, A., et al.: Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models, Journal of Climate, 28, 952–977, 2015.

Guba, O., Taylor, M., Ullrich, P., Overfelt, J., and Levy, M.: The spectral element method (SEM) on variable-resolution grids: Evaluating grid sensitivity and resolution-aware numerical viscosity, Geoscientific Model Development, 7, 2803–2816, 2014.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hagos, S., Leung, R., Rauscher, S. A., and Ringler, T.: Error Characteristics of Two Grid Refinement Approaches in Aquaplanet Simulations: MPAS-A and WRF, Monthly Weather Review, 141, 3022 – 3036, https://doi.org/10.1175/MWR-D-12-00338.1, 2013.

Hannah, W. M., Bradley, A. M., Guba, O., Tang, Q., Golaz, J.-C., and Wolfe, W.: Separating Physics and Dynamics grids for Improved Computational Efficiency in Spectral Element Earth System Models, J. Adv. Model Earth Sy., 13, e2020MS002 419, https://doi.org/10.1029/2020ms002419, 2021.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: GLOBAL SURFACE TEMPERATURE CHANGE, Reviews of Geophysics, 48, https://doi.org/10.1029/2010RG000345, 2010.

Harris, L. M. and Lin, S.-J.: A Two-Way Nested Global-Regional Dynamical Core on the Cubed-Sphere Grid, Monthly Weather Review, 141, 283 – 306, https://doi.org/10.1175/MWR-D-11-00201.1, 2013.

Harris, L. M. and Lin, S.-J.: Global-to-Regional Nested Grid Climate Simulations in the GFDL High Resolution Atmospheric Model, Journal of Climate, 27, 4890 – 4910, https://doi.org/10.1175/JCLI-D-13-00596.1, 2014.

Hazelton, A. T., Harris, L., and Lin, S.-J.: Evaluation of Tropical Cyclone Structure Forecasts in a High-Resolution Version of the Multiscale GFDL fvGFS Model, Weather and Forecasting, 33, 419 – 442, https://doi.org/10.1175/WAF-D-17-0140.1, 2018.

Held, I. M. and Shell, K. M.: Using relative humidity as a state variable in climate feedback analysis, Journal of Climate, 25, 2578–2582, 2012.

905 Herrington, A. R., Lauritzen, P. H., Lofverstrom, M., Lipscomb, W. H., Gettelman, A., and Taylor, M. A.: Impact of grids and dynamical cores in CESM2.2 on the surface mass balance of the Greenland Ice Sheet, Journal of Advances in Modeling Earth Systems, https://doi.org/10.1029/2022ms003192, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee,
910 D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu,
915 F., Jankowiak, I., and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol characterization, Remote Sensing of Environment, 66, 1–16, https://doi.org/10.1016/s0034-4257(98)00031-5, 1998.

Huang, X. and Ullrich, P. A.: Irrigation impacts on California's climate with the variable-resolution CESM, Journal of Advances in Modeling Earth Systems, 8, 1151–1163, https://doi.org/10.1002/2016MS000656, 2016.

Huang, X. and Ullrich, P. A.: The Changing Character of Twenty-First-Century Precipitation over the Western United States in the Variable-
920 Resolution CESM, Journal of Climate, 30, 7555 – 7575, https://doi.org/10.1175/JCLI-D-16-0673.1, 2017.

Huang, X., Rhoades, A. M., Ullrich, P. A., and Zarzycki, C. M.: An evaluation of the variable-resolution CESM for modeling California's climate, Journal of Advances in Modeling Earth Systems, 8, 345–369, https://doi.org/10.1002/2015MS000559, 2016.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of
925 hydrometeorology, 8, 38–55, 2007.

Hughes, O. K. and Jablonowski, C.: A Mountain-Induced Moist Baroclinic Wave Test Case for the Dynamical Cores of Atmospheric General Circulation Models, EGUsphere, preprint, 2023.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., Lawrence,
930 P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F. N., van Vuuren, D. P., and Zhang, X.: Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6, Geoscientific Model Development, 13, 5425–5464, https://doi.org/10.5194/gmd-13-5425-2020, 2020.

Jiang, X., Lau, N.-C., and Klein, S. A.: Role of eastward propagating convection systems in the diurnal cycle and seasonal mean of summertime rainfall over the U.S. Great Plains, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL027022, 2006.

935 Jung, T., Gulev, S. K., Rudeva, I., and Soloviov, V.: Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model, Quarterly Journal of Royal Meteorological Society, 132, 1839–1857, https://doi.org/10.1256/qj.05.212, 2006.

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

and Claussen, M.: The ICON Earth System Model Version 1.0, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 813, https://doi.org/10.1029/2021MS002813, 2022.

Kapnick, S. B., Yang, X., Vecchi, G. A., Delworth, T. L., Gudgel, R., Malyshev, S., Milly, P. C. D., Shevliakova, E., Underwood, S., and Margulis, S. A.: Potential for western US seasonal snowpack prediction, Proceedings of the National Academy of Sciences, 115, 1180–1185, https://doi.org/10.1073/pnas.1716760115, 2018.

Larson, V. E.: CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere, https://doi.org/10.48550/ARXIV.1711.03675, 2017.

Lauritzen, P. H., Bacmeister, J. T., Callaghan, P. F., and Taylor, M. A.: NCAR_Topo (v1.0): NCAR global model topography generation software for unstructured grids, Geoscientific Model Development, 8, 3975–3986, https://doi.org/10.5194/gmd-8-3975-2015, 2015.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., et al.: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245–4287, 2019.

Leung, L. R., Ringler, T., Collins, W. D., Taylor, M., and Ashfaq, M.: A Hierarchical Evaluation of Regional Climate Simulations, Eos, Transactions American Geophysical Union, 94, 297–298, https://doi.org/10.1002/2013EO340001, 2013.

Leung, L. R., Bader, D. C., Taylor, M. A., and McCoy, R. B.: An Introduction to the E3SM Special Collection: Goals, Science Drivers, Development, and Analysis, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 821, https://doi.org/10.1029/2019MS001821, e2019MS001821, 2020.

Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., and Leung, L. R.: A physically based runoff routing model for land surface and earth system models, Journal of Hydrometeorology, 14, 808–828, 2013.

Li, H.-Y., Leung, L. R., Getirana, A., Huang, M., Wu, H., Xu, Y., Guo, J., and Voisin, N.: Evaluating global streamflow simulations by a physically based routing model coupled with the community land model, Journal of Hydrometeorology, 16, 948–971, 2015a.

Li, H.-Y., Ruby Leung, L., Tesfa, T., Voisin, N., Hejazi, M., Liu, L., Liu, Y., Rice, J., Wu, H., and Yang, X.: Modeling stream temperature in the A nthropocene: An earth system modeling approach, Journal of Advances in Modeling Earth Systems, 7, 1661–1679, 2015b.

Liang, Y., Yang, B., Wang, M., Tang, J., Sakaguchi, K., Leung, L. R., and Xu, X.: Multiscale Simulation of Precipitation Over East Asia by Variable Resolution CAM-MPAS, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 656, https://doi.org/10.1029/2021MS002656, e2021MS002656 2021MS002656, 2021.

Liu, Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., Ghan, S. J., and Rasch, P. J.: Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model, Geoscientific Model Development, 9, 505–522, https://doi.org/10.5194/gmd-9-505-2016, 2016.

Liu, W., Ullrich, P., Li, J., Zarzycki, C. M., Caldwell, P. M., Leung, L. R., and Qian, Y.: The June 2012 North American Derecho: A testbed for evaluating regional and global climate modeling systems at cloud-resolving scales, Earth and Space Science Open Archive, p. 39, https://doi.org/10.1002/essoar.10511614.1, 2022.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: 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, 895–918, https://doi.org/10.1175/jcli-d-17-0208.1, 2018.

Melillo, J. M., Richmond, T. C., and Yohe, G. W. E.: Climate Change Impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program, U.S. Global Change Research Program, p. 841, https://doi.org/10.7930/J0Z31WJ2, 2014.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD017187, 2012.

Palazzi, E., Mortarini, L., Terzago, S., and Von Hardenberg, J.: Elevation-dependent warming in global climate model simulations at high spatial resolution, Climate Dynamics, 52, 2685–2702, https://doi.org/10.1007/s00382-018-4287-z, 2019.

Rahimi, S. R., Wu, C., Liu, X., and Brown, H.: Exploring a Variable-Resolution Approach for Simulating Regional Climate Over the Tibetan Plateau Using VR-CESM, Journal of Geophysical Research: Atmospheres, 124, 4490–4513, https://doi.org/10.1029/2018JD028925, 2019.

Rauscher, S. A., Ringler, T. D., Skamarock, W. C., and Mirin, A. A.: Exploring a Global Multiresolution Modeling Approach Using Aqua-planet Simulations, Journal of Climate, 26, 2432 – 2452, https://doi.org/10.1175/JCLI-D-12-00154.1, 2013.

Reed, K. A. and Jablonowski, C.: Idealized tropical cyclone simulations of intermediate complexity: a test case for AGCMs, J. Adv. Model. Earth Syst., 4, M04 001, 2012.

Reed, K. A., Wehner, M. F., and Zarzycki, C. M.: Attribution of 2020 hurricane season extreme rainfall to human-induced climate change, Nature communications, 13, 1–6, https://doi.org/10.1038/s41467-022-29379-1, 2022.

Rhoades, A. M., Huang, X., Ullrich, P. A., and Zarzycki, C. M.: Characterizing Sierra Nevada Snowpack Using Variable-Resolution CESM, Journal of Applied Meteorology and Climatology, 55, 173 – 196, https://doi.org/10.1175/JAMC-D-15-0156.1, 2016.

Rhoades, A. M., Ullrich, P. A., and Zarzycki, C. M.: Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM, Climate Dynamics, 50, 261–288, https://doi.org/10.1007/s00382-017-3606-0, 2017.

Rhoades, A. M., Jones, A. D., and Ullrich, P. A.: Assessing Mountains as Natural Reservoirs With a Multimetric Framework, Earth's Future, 6, 1221–1241, https://doi.org/10.1002/2017EF000789, 2018a.

Rhoades, A. M., Jones, A. D., and Ullrich, P. A.: The Changing Character of the California Sierra Nevada as a Natural Reservoir, Geophysical Research Letters, 45, 13,008–13,019, https://doi.org/10.1029/2018GL080308, 2018b.

Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Johansen, H., Margulis, S. A., Morrison, H., Xu, Z., and Collins, W. D.: Sensitivity of Mountain Hydroclimate Simulations in Variable-Resolution CESM to Microphysics and Horizontal Resolution, Journal of Advances in Modeling Earth Systems, 10, 1357–1380, https://doi.org/10.1029/2018MS001326, 2018c.

Rhoades, A. M., Jones, A. D., O'Brien, T. A., O'Brien, J. P., Ullrich, P. A., and Zarzycki, C. M.: Influences of North Pacific Ocean Domain Extent on the Western U.S. Winter Hydroclimatology in Variable-Resolution CESM, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 977, https://doi.org/10.1029/2019JD031977, e2019JD031977 10.1029/2019JD031977, 2020a.

Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., Ullrich, P. A., Wehner, M., and Zhou, Y.: The Shifting Scales of Western U.S. Landfalling Atmospheric Rivers Under Climate Change, Geophysical Research Letters, 47, e2020GL089 096, https://doi.org/10.1029/2020GL089096, 2020b.

Rhoades, A. M., Hatchett, B. J., Risser, M. D., Collins, W. D., Bambach, N. E., Huning, L. S., McCrary, R., Siirila-Woodburn, E. R., Ullrich, P. A., Wehner, M. F., Zarzycki, C. M., and Jones, A. D.: Asymmetric emergence of low-to-no snow in the midlatitudes of the American Cordillera, Nature Climate Change, 12, 1151–1159, https://doi.org/10.1038/s41558-022-01518-y, 2022.

Richter, J. H., Sassi, F., and Garcia, R. R.: Toward a Physically Based Gravity Wave Source Parameterization in a General Circulation Model, Journal of the Atmospheric Sciences, 67, 136 – 156, https://doi.org/10.1175/2009JAS3112.1, 2010.

Richter, J. H., Chen, C.-C., Tang, Q., Xie, S., and Rasch, P. J.: Improved Simulation of the QBO in E3SMv1, Journal of Advances in Modeling Earth Systems, 11, 3403–3418, https://doi.org/10.1029/2019MS001763, 2019.

Riley, G. T., Landin, M. G., and Bosart, L. F.: The Diurnal Variability of Precipitation across the Central Rockies and Adjacent Great Plains, Monthly Weather Review, 115, 1161 – 1172, https://doi.org/10.1175/1520-0493(1987)115<1161:TDVOPA>2.0.CO;2, 1987.

Ringler, T., Ju, L., and Gunzburger, M.: A multiresolution method for climate system modeling: application of spherical centroidal Voronoi tessellations, Ocean Dynamics, 58, 475–498, https://doi.org/10.1007/s10236-008-0157-2, 2008.

1020 Running, S., Mu, Q., and Zhao, M.: MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006, Tech. rep., https://doi.org/10.5067/MODIS/MOD16A2.006, 2017.

Sakaguchi, K., Leung, L. R., Zhao, C., Yang, Q., Lu, J., Hagos, S., Rauscher, S. A., Dong, L., Ringler, T. D., and Lauritzen, P. H.: Exploring a Multiresolution Approach Using AMIP Simulations, Journal of Climate, 28, 5549 – 5574, https://doi.org/10.1175/JCLI-D-14-00729.1, 2015.

1025 Sakaguchi, K., Lu, J., Leung, L. R., Zhao, C., Li, Y., and Hagos, S.: Sources and pathways of the upscale effects on the Southern Hemisphere jet in MPAS-CAM4 variable-resolution simulations, Journal of Advances in Modeling Earth Systems, 8, 1786–1805, https://doi.org/10.1002/2016MS000743, 2016.

Santanello, J. A., Peters-Lidard, C. D., Kumar, S. V., Alonge, C., and Tao, W.-K.: A Modeling and Observational Framework for Diagnosing Local Land–Atmosphere Coupling on Diurnal Time Scales, Journal of Hydrometeorology, 10, 577 – 599,
1030 https://doi.org/10.1175/2009JHM1066.1, 2009.

Santanello, J. A., Dirmeyer, P. A., Ferguson, C. R., Findell, K. L., Tawfik, A. B., Berg, A., Ek, M., Gentine, P., Guillod, B. P., van Heerwaarden, C., Roundy, J., and Wulfmeyer, V.: Land–Atmosphere Interactions: The LoCo Perspective, Bulletin of the American Meteorological Society, 99, 1253 – 1272, https://doi.org/10.1175/BAMS-D-17-0001.1, 2018.

Siirila-Woodburn, E., Rhoades, A. M., Hatchett, B. J., Huning, L., Szinai, J., Tague, C., Nico, P. S., Feldman, D., Jones, A. D., Collins,
1035 W. D., and Kaatz, L.: A low-to-no snow future and its impacts on water resources in the western United States, Nature Reviews Earth and Environment, 2, 800–819, https://doi.org/10.1038/s43017-021-00219-y, 2021.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), Journal of Climate, 21, 2283 – 2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., and Shields, C. A.: Quantifying climate feedbacks using radiative kernels,
1040 Journal of Climate, 21, 3504–3520, 2008.

Sun, J., Zhang, K., Wan, H., Ma, P.-L., Tang, Q., and Zhang, S.: Impact of Nudging Strategy on the Climate Representativeness and Hindcast Skill of Constrained EAMv1 Simulations, Journal of Advances in Modeling Earth Systems, 11, 3911–3933, https://doi.org/10.1029/2019MS001831, 2019.

Tang, Q., Klein, S. A., Xie, S., Lin, W., Golaz, J.-C., Roesler, E. L., Taylor, M. A., Rasch, P. J., Bader, D. C., Berg, L. K., Caldwell, P.,
1045 Giangrande, S. E., Neale, R. B., Qian, Y., Riihimaki, L. D., Zender, C. S., Zhang, Y., and Zheng, X.: Regionally refined test bed in E3SM atmosphere model version 1 (EAMv1) and applications for high-resolution modeling, Geoscientific Model Development, 12, 2679–2706, https://doi.org/10.5194/gmd-12-2679-2019, 2019.

Tang, Q., Prather, M. J., Hsu, J., Ruiz, D. J., Cameron-Smith, P. J., Xie, S., and Golaz, J.-C.: Evaluation of the interactive stratospheric ozone (O3v2) module in the E3SM version 1 Earth system model, Geoscientific Model Development, 14, 1219–1236,
1050 https://doi.org/10.5194/gmd-14-1219-2021, 2021.

Tao, C., Zhang, Y., Tang, Q., Ma, H.-Y., Ghate, V. P., Tang, S., Xie, S., and Santanello, J. A.: Land–Atmosphere Coupling at the U.S. Southern Great Plains: A Comparison on Local Convective Regimes between ARM Observations, Reanalysis, and Climate Model Simulations, Journal of Hydrometeorology, 22, 463 – 481, https://doi.org/10.1175/JHM-D-20-0078.1, 2021.

Taylor, M. A. and Fournier, A.: A compatible and conservative spectral element method on unstructured grids, J. Comput. Phys., 229, 5879–5895, https://doi.org/10.1016/j.jcp.2010.04.008, 2010.

Taylor, M. A., Guba, O., Steyer, A., Ullrich, P. A., Hall, D. M., and Eldred, C.: An Energy Consistent Discretization of the Nonhydrostatic Equations in Primitive Variables, Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019MS001783, 2020.

Teixeira, J., Cardoso, S., Bonazzola, M., Cole, J., DelGenio, A., DeMott, C., Franklin, C., Hannay, C., Jakob, C., Jiao, Y., et al.: Tropical and subtropical cloud transitions in weather and climate prediction models: The GCSS/WGNE Pacific Cross-Section Intercomparison (GPCI), Journal of Climate, 24, 5223–5256, 2011.

Ulbrich, U., Leckebusch, G. C., and Pinto, J. G.: Extra-tropical cyclones in the present and future climate: a review, Theoretical and applied climatology, 96, 117–131, https://doi.org/10.1007/s00704-008-0083-8, 2009.

Ullrich, P.: Spherical Quadrilateral Grid Generator (SQuadGen), https://doi.org/10.5281/zenodo.7102543, v1.1, 2022.

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

Ullrich, P. A. and Zarzycki, C. M.: TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids, Geoscientific Model Development, 10, 1069–1090, https://doi.org/10.5194/gmd-10-1069-2017, 2017.

Ullrich, P. A., Melvin, T., Jablonowski, C., and Staniforth, A.: A proposed baroclinic wave test case for deep and shallow-atmosphere dynamical cores, Quart. J. Roy. Meteor. Soc., 140, 1590–1602, 2014.

Ullrich, P. A., Devendran, D., and Johansen, H.: Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part II, Monthly Weather Review, 144, 1529–1549, 2016.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets, Geoscientific Model Development, 14, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021, 2021.

van Kampenhout, L., Rhoades, A. M., Herrington, A. R., Zarzycki, C. M., Lenaerts, J. T. M., Sacks, W. J., and van den Broeke, M. R.: Regional grid refinement in an Earth system model: impacts on the simulated Greenland surface mass balance, The Cryosphere, 13, 1547–1564, https://doi.org/10.5194/tc-13-1547-2019, 2019.

Van Roekel, L. P., Golaz, J.-C., and et al.: E3SMv2 NARRM ocean/sea ice overview, Journal of Advances in Modeling Earth Systems, to be submitted, 2023.

Veneziani, M., Maslowski, W., Lee, Y. J., D'Angelo, G., Osinski, R., Petersen, M. R., Weijer, W., Craig, A. P., Wolfe, J. D., Comeau, D., and Turner, A. K.: An evaluation of the E3SMv1 Arctic ocean and sea-ice regionally refined model, Geoscientific Model Development, 15, 3133–3160, https://doi.org/10.5194/gmd-15-3133-2022, 2022.

Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., Ganguly, D., Rasch, P. J., Burrows, S. M., Ghan, S. J., Lou, S., Qian, Y., Yang, Y., Feng, Y., Flanner, M., Leung, L. R., Liu, X., Shrivastava, M., Sun, J., Tang, Q., Xie, S., and Yoon, J.-H.: Aerosols in the E3SM Version 1: New Developments and Their Impacts on Radiative Forcing, Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019MS001851, 2020.

Wu, C., Liu, X., Lin, Z., Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Lu, Z., and Rahimi-Esfarjani, S. R.: Exploring a Variable-Resolution Approach for Simulating Regional Climate in the Rocky Mountain Region Using the VR-CESM, Journal of Geophysical Research: Atmospheres, 122, 10,939–10,965, https://doi.org/10.1002/2017JD027008, 2017.

Xie, S., Lin, W., Rasch, P. J., Ma, P.-L., Neale, R., Larson, V. E., Qian, Y., Bogenschutz, P. A., Caldwell, P., Cameron-Smith, P., Golaz, J.-C., Mahajan, S., Singh, B., Tang, Q., Wang, H., Yoon, J.-H., Zhang, K., and Zhang, Y.: Understanding Cloud and Convec-

tive Characteristics in Version 1 of the E3SM Atmosphere Model, Journal of Advances in Modeling Earth Systems, 10, 2618–2644, https://doi.org/10.1029/2018MS001350, 2018.

Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S., Zheng, X., Golaz, J.-C., Zhang, G. J., and Zhang, M.: Improved Diurnal Cycle of Precipitation in E3SM With a Revised Convective Triggering Function, Journal of Advances in Modeling Earth Systems, 11, 2290–2310, https://doi.org/10.1029/2019MS001702, 2019.

Xu, Z. and Di Vittorio, A.: Hydrological analysis in watersheds with a variable-resolution global climate model (VR-CESM), Journal of Hydrology, 601, 126 646, https://doi.org/10.1016/j.jhydrol.2021.126646, 2021.

Xu, Z., Di Vittorio, A., Zhang, J., Rhoades, A., Xin, X., Xu, H., and Xiao, C.: Evaluating Variable-Resolution CESM Over China and Western United States for Use in Water-Energy Nexus and Impacts Modeling, Journal of Geophysical Research: Atmospheres, 126, e2020JD034 361, https://doi.org/10.1029/2020JD034361, e2020JD034361 2020JD034361, 2021.

Xu, Z., Chang, A., and Di Vittorio, A.: Evaluating and projecting of climate extremes using a variable-resolution global climate model (VR-CESM), Weather and Climate Extremes, 38, 100 496, https://doi.org/10.1016/j.wace.2022.100496, 2022.

Zarzycki, C. M. and Jablonowski, C.: A multidecadal simulation of Atlantic tropical cyclones using a variable-resolution global atmospheric general circulation model, Journal of Advances in Modeling Earth Systems, 6, 805–828, https://doi.org/10.1002/2014MS000352, 2014.

Zarzycki, C. M. and Jablonowski, C.: Experimental Tropical Cyclone Forecasts Using a Variable-Resolution Global Model, Monthly Weather Review, 143, 4012 – 4037, https://doi.org/10.1175/MWR-D-15-0159.1, 2015.

Zarzycki, C. M., Levy, M. N., Jablonowski, C., Overfelt, J. R., Taylor, M. A., and Ullrich, P. A.: Aquaplanet Experiments Using CAM's Variable-Resolution Dynamical Core, Journal of Climate, 27, 5481 – 5503, https://doi.org/10.1175/JCLI-D-14-00004.1, 2014.

Zarzycki, C. M., Jablonowski, C., Thatcher, D. R., and Taylor, M. A.: Effects of Localized Grid Refinement on the General Circulation and Climatology in the Community Atmosphere Model, Journal of Climate, 28, 2777 – 2803, https://doi.org/10.1175/JCLI-D-14-00599.1, 2015.

Zarzycki, C. M., Ullrich, P. A., and Reed, K. A.: Metrics for Evaluating Tropical Cyclones in Climate Data, Journal of Applied Meteorology and Climatology, 60, 643 – 660, https://doi.org/10.1175/JAMC-D-20-0149.1, 2021.

Zender, C. S.: Analysis of self-describing gridded geoscience data with netCDF Operators (NCO), Environmental Modelling & Software, 23, 1338–1342, 2008.

Zhang, C., Golaz, J.-C., Forsyth, R., Vo, T., Xie, S., Shaheen, Z., Potter, G. L., Asay-Davis, X. S., Zender, C. S., Lin, W., Chen, C.-C., Terai, C. R., Mahajan, S., Zhou, T., Balaguru, K., Tang, Q., Tao, C., Zhang, Y., Emmenegger, T., and Ullrich, P.: The E3SM Diagnostics Package (E3SM Diags v2.6): A Python-based Diagnostics Package for Earth System Models Evaluation, Geoscientific Model Development Discussions, 2022, 1–35, https://doi.org/10.5194/gmd-2022-38, 2022a.

Zhang, G. J. and McFarlane, N. A.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model, Atmosphere-Ocean, 33, 407–446, https://doi.org/10.1080/07055900.1995.9649539, 1995.

Zhang, H.-M., Huang, B., Lawrimore, J., Menne, M., and Smith, T. M.: NOAA Global Surface Temperature Dataset (NOAAGlobalTemp), Version 4.0, https://doi.org/10.7289/V5FN144H, 2015.

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P.-L., Rasch, P. J., Neubauer, D., and Lohmann, U.: Technical Note: On the use of nudging for aerosol–climate model intercomparison studies, Atmospheric Chemistry and Physics, 14, 8631–8645, https://doi.org/10.5194/acp-14-8631-2014, 2014.

Zhang, K., Rasch, P. J., Taylor, M. A., Wan, H., Leung, R., Ma, P.-L., Golaz, J.-C., Wolfe, J., Lin, W., Singh, B., Burrows, S., Yoon, J.-H., Wang, H., Qian, Y., Tang, Q., Caldwell, P., and Xie, S.: Impact of numerical choices on water conservation in the E3SM Atmosphere Model version 1 (EAMv1), Geoscientific Model Development, 11, 1971–1988, https://doi.org/10.5194/gmd-11-1971-2018, 2018.

Zhang, K., Zhang, W., Wan, H., Rasch, P. J., Ghan, S. J., Easter, R. C., Shi, X., Wang, Y., Wang, H., Ma, P.-L., Zhang, S., Sun, J., Burrows, S. M., Shrivastava, M., Singh, B., Qian, Y., Liu, X., Golaz, J.-C., Tang, Q., Zheng, X., Xie, S., Lin, W., Feng, Y., Wang, M., Yoon, J.-H., and Leung, L. R.: Effective radiative forcing of anthropogenic aerosols in E3SM version 1: historical changes, causality, decomposition, and parameterization sensitivities, Atmospheric Chemistry and Physics, 22, 9129–9160, https://doi.org/10.5194/acp-22-9129-2022, 2022b.

Zhang, S., Zhang, K., Wan, H., and Sun, J.: Further improvement and evaluation of nudging in the E3SM Atmosphere Model version 1 (EAMv1): simulations of the mean climate, weather events, and anthropogenic aerosol effects, Geoscientific Model Development, 15, 6787–6816, https://doi.org/10.5194/gmd-15-6787-2022, 2022c.

Zhang, Y. and et al.: Understanding Changes in Cloud Simulations from E3SM Version 1 to Version 2, Geoscientific Model Development, to be submitted, 2023.

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., Rasch, P. J., Qian, Y., Tang, Q., and Ma, H.-Y.: Evaluation of Clouds in Version 1 of the E3SM Atmosphere Model With Satellite Simulators, Journal of Advances in Modeling Earth Systems, 11, 1253–1268, https://doi.org/10.1029/2018MS001562, 2019.

Zhou, T., Leung, L. R., Leng, G., Voisin, N., Li, H.-Y., Craig, A. P., Tesfa, T., and Mao, Y.: Global irrigation characteristics and effects simulated by fully coupled land surface, river, and water management models in E3SM, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 069, 2020.