

# Separating Physics and Dynamics grids for Improved Computational Efficiency in Spectral Element Earth System Models

Walter M. Hannah<sup>1</sup>, Andrew M. Bradley<sup>2</sup>, Oksana Guba<sup>2</sup>, Qi Tang<sup>1</sup>, Jean-Christophe Golaz<sup>1</sup>, and Jon Wolfe<sup>3</sup>

<sup>1</sup>*Lawrence Livermore National Laboratory, Livermore, CA*

<sup>2</sup>*Sandia National Laboratories, Albuquerque, NM*

<sup>3</sup>*Los Alamos National Laboratory, Los Alamos, NM*

## Key Points:

- A method is presented for defining a finite volume physics grid in a spectral element model that allows for regional refinement.
- The new method is shown to qualitatively preserve the model solution and effective resolution.
- A relatively coarse physics grid increases the speed of physics by roughly 60% to 120% depending on how the computational resources are configured.

**Abstract:** Previous studies have shown that atmospheric models with a spectral element grid can benefit from putting physics calculations on a relatively coarse finite volume grid. Here we demonstrate an alternative high-order, element-based mapping approach used to implement a quasi-equal-area, finite volume physics grid in E3SM. Unlike similar methods, the new method in E3SM requires topology data purely local to each spectral element, which trivially allows for regional mesh refinement. Simulations with physics grids defined by  $2 \times 2$ ,  $3 \times 3$ , and  $4 \times 4$  divisions of each element are shown to verify that the alternative physics grid does not qualitatively alter the model solution. The model performance is substantially affected by the reduction of physics columns when using the  $2 \times 2$  grid, which can increase the throughput of physics calculations by roughly 60% to 120% depending on whether the computational resources are configured to maximize throughput or efficiency. A pair of regionally refined cases are also shown to highlight the refinement capability.

**Plain Language Summary:** Most atmospheric models use the same grid for dynamics (e.g. advection) and physics (e.g. clouds). For spectral element models the grid uses irregularly spaced points and the treatment of element edges can lead to grid imprinting bias. Previous studies have shown that using a regularly spaced physics grid in a spectral element model can alleviate the grid imprinting biases. This alternative physics grid can also reduce the computational cost of the model if the physics grid is coarser than the dynamics grid. This study presents a new approach for using a regularly spaced physics grid in a global spectral element model that additionally allows mesh refinement for regionally focused simulations. The use of a relatively coarse physics grid is shown to make the model faster without qualitatively degrading the simulated climate.

## 1 Introduction

Global atmospheric models traditionally evaluate physical process parameterizations on the same grid as the dynamical core. However, with certain element-based Galerkin methods this co-located

approach can lead to grid imprinting biases [10, 8] that arise when the derivatives of the global basis set are not continuous (i.e. they are  $C^0$ ). In contrast to finite volume methods, element-based Galerkin methods use irregularly spaced quadrature points within each element to define the grid. When physics tendencies are computed on these quadrature points, discontinuous derivatives can occur on element edges and produce unrealistic localized extrema that drive the grid imprinting biases [10].

Recently, [9] demonstrated that a relatively coarse quasi-equal-area finite volume grid for physics parameterizations with the NCAR CAM spectral element (CAM-SE) dynamical core [5] can produce a qualitatively identical solution to the default configuration with a  $5/9$  reduction in the number of physics columns. This approach eliminates the grid imprinting problem as well as some spurious noise in the vertical velocity field over steep topography. Additionally, the effective resolution of spectral element dynamics is coarser than the average spacing of the nodal points [12], so considering physics processes at this coarser scale makes physical sense and reduces the computational burden of the model.

The quasi-equal-area finite volume physics grid (hereafter referred to as “physgrid”) method presented by [10] was partly motivated by the need to support a finite volume semi-Lagrangian tracer transport scheme, known as CSLAM [16]. CSLAM uses tensor-cubic interpolation to map data from the FV physics grid to the dynamics grid and requires a stencil that extends beyond the element boundaries. [10] use a similar cubic interpolation that is smooth across element boundaries for their physgrid implementation. This is advantageous since the boundary exchange with the surface component models is on the same grid, but also carries the disadvantage that regional mesh refinement is much more complicated.

An alternative implementation of a finite volume physgrid was recently developed for use in the Energy Exascale Earth System Model (E3SM) [7], hereafter G19, that is designed to support regional mesh refinement of quasi-uniform grids. The part of the high-order finite-volume reconstruction that depends on a structured grid is local to a spectral element, and thus is independent of the topology of the spectral element grid.

The goal of this paper is to describe the physgrid implementation in the atmospheric component of E3SM and verify that it does not qualitatively affect the model behavior. The mapping technique is detailed in Section 2. Model experiments are described in Section 3 followed by analysis of the results in Section 4. Conclusions are discussed in Section 5.

## 2 High-order, Property-preserving Remap Operators

The E3SM atmospheric dynamical core is known as the High-Order Methods Modeling Environment (HOMME). The dynamics grid is built upon unstructured quadrilaterals of a cubed-sphere mesh with  $n_p \times n_p$  Gauss-Lobatto-Legendre (GLL) quadrature points located at the roots of the basis set [24, 25, 5]. A quasi-uniform cubed-sphere mesh is specified by the parameter  $n_e$ ; this mesh has  $n_e \times n_e$  spectral elements per cube face. A regionally refined mesh (RRM) retains the overall cubed-sphere structure but refines a region of the global domain to produce a locally fully unstructured mesh. While E3SM can be configured to run with any value of  $n_p$ , all configurations use  $n_p = 4$ , for which the basis functions are third-order polynomials (e.g. ne30np4 for the standard resolution configuration of E3SMv1). The evolution of the solution is communicated among elements via direct stiffness summation (DSS) [4] to form the global basis set, which leads to discontinuous derivatives at element boundaries (i.e. the solution is  $C^0$ ).

The new physics grid divides each quadrilateral element into  $n_f \times n_f$  finite-volume (FV) cells that are uniform on the element’s reference coordinate and approximately uniform after mapping to the sphere. Since E3SM always uses  $n_p = 4$ , in the case of a quasi-uniform cubed sphere mesh, an

atmosphere grid set is specified with just  $n_e$  and  $n_f$  (e.g. ne30pg2). The default physics configuration, in which physics columns exist on the GLL nodes, continues to be referred to as “np4” (e.g. ne30np4).

This section describes our physgrid remap algorithms.

## 2.1 Definitions

The dynamics and physics mesh parameters are  $(n_p, n_f)$ . In this work, we assume  $n_f \leq n_p$ . An *element* is a spectral element, and a *subcell* is a physics finite volume (FV) cell inside the element. Figure 1 illustrates an example of a spectral element. The blue square outlines one element, and the green filled circles mark GLL points. Each of the four red-dashed squares in Figure 1 is a subcell. We use “finite volume” in only a loose sense; the discretization shares with a fully FV discretization the characteristics of a constant basis function over a subcell and a high-order reconstruction combining data from multiple subcells, but there is no flux function as in a fully FV method. An element has reference domain  $R$ ; this domain is the full element outlined in blue in Figure 1. Subcell  $c$  in an element has reference domain  $R_c^f \subset R$ ; in Figure 1, each red square has its own reference domain  $R_c^f$ . A reference element  $R$  is subdivided uniformly into  $n_f^2$  squares  $R_c^f$ . We use  $\mathbf{r} \in R$  to refer to a reference point, and similarly for  $R_c^f$ .

Unless otherwise indicated, sums are over quantities within a single element. A bold-face quantity is a vector of coefficients in an element. A superscript letter denotes the basis; for example,  $\rho^f$  are the FV-basis coefficients for an approximation of  $\rho$ , and  $\rho^p$  are the GLL-basis coefficients.

$b_i^p(\mathbf{r})$  is a GLL tensor-product basis function local to an element, of which there are  $n_p^2$  in an element. The function corresponding to a vector  $\mathbf{p}$  is  $\sum_i p_i b_i^p(\mathbf{r})$ , where  $\mathbf{p}$  provides nodal values at nodes  $\mathbf{r}_i^p$ .  $\mathcal{I}^{p' \rightarrow p}$  projects  $\mathbf{p}'$ , a quantity in the GLL  $n_{p'}$ -basis, to  $\mathbf{p}$ , a quantity in the GLL  $n_p$ -basis, using the  $n_{p'}$ -basis functions to interpolate  $\mathbf{p}'$  to the  $n_p$ -basis nodes. Row  $j$  of  $\mathcal{I}^{p' \rightarrow p}$  provides  $p_j = \sum_i p'_i b_i^{p'}(\mathbf{r}_j^p)$ . In this work we have  $n_{p'} \leq n_p$ , and thus this projection has no error.

The weight associated with a GLL basis function is  $w_i^p \equiv \int_R b_i(\mathbf{r}) \, d\mathbf{r}$ . Similarly, the weight associated with the FV basis function, the constant function over a subcell, is  $w_c^f \equiv \int_{R_c^f} d\mathbf{r}$ .

$J$  is the Jacobian determinant of the map from reference element to sphere.  $J^p$  includes a constant factor  $\alpha$  that is determined so that the sum of  $J_i^p w_i^p$  over the sphere is  $4\pi$ .  $J^f$  will be discussed after developing the first remap operator.

## 2.2 Linear, element-local, remap operators

In this subsection, we describe linear, element-local, remap operators. This subsection and subsection 2.3 focus on the operations performed within a single element; subsequent subsections then describe the operations on the full grids.

### 2.2.1 GLL to FV

It is natural to map GLL data to FV data by averaging the GLL density over the FV subcell:

$$\rho_c^f = \left( \int_{R_c^f} J^p(\mathbf{r}) \, d\mathbf{r} \right)^{-1} \int_{R_c^f} \rho^p(\mathbf{r}) J^p(\mathbf{r}) \, d\mathbf{r} \approx (J_c^f w_c^f)^{-1} \sum_i \rho_i^p J_i^p M_{ci}^p, \quad (1)$$

where  $M_{ci}^p \equiv \int_{R_c^f} b_i(\mathbf{r}) \, d\mathbf{r}$ , the approximation is the spectral element quadrature approximation, and  $J_c^f$  will be defined shortly. In HOMME, it is best to think of spectral element operators as being applied to the combined quantity  $\text{diag}(\mathbf{J})\boldsymbol{\rho}$ , where  $\text{diag}(\cdot)$  is the diagonal matrix having its argument

on the diagonal. Then the linear operator to map GLL reference-element data to FV reference-subcell data is

$$A^{p \rightarrow f} \equiv \text{diag}(\mathbf{w}^f)^{-1} M^p. \quad (2)$$

With this operator we can write (1) as  $\boldsymbol{\rho}^f = \text{diag}(\mathbf{J}^f)^{-1} A^{p \rightarrow f} \text{diag}(\mathbf{J}^p) \boldsymbol{\rho}^p$ , where  $\mathbf{J}^f \equiv A^{p \rightarrow f} \mathbf{J}^p$ . For convenience, let  $B^{p \rightarrow f} \equiv \text{diag}(\mathbf{J}^f)^{-1} A^{p \rightarrow f} \text{diag}(\mathbf{J}^p)$ , so that we can write  $\boldsymbol{\rho}^f = B^{p \rightarrow f} \boldsymbol{\rho}^p$ .

### 2.2.2 FV to GLL

Our maps operate on the reference element; thus, in what follows,  $\mathbf{J}$  does not appear and  $\mathbf{f}$  and  $\mathbf{p}$  are quantities on the reference element. To determine the map from FV to GLL data in an element,  $A^{f \rightarrow p}$ , we make the following design requirements.

R1. If  $\mathbf{p} = A^{f \rightarrow p} \mathbf{f}$ , then  $A^{p \rightarrow f} \mathbf{p} = \mathbf{f}$ .

R2. If (a)  $\mathbf{f} = A^{p \rightarrow f} \mathbf{p}$ , (b)  $n_{p'} = n_f$ , and (c)  $\mathbf{p} = \mathcal{I}^{p' \rightarrow p} \mathbf{p}'$ , then (d)  $A^{f \rightarrow p} \mathbf{f} = \mathbf{p}$ .

Each requirement specifies a limited idempotence. Requirement R1 specifies that repeated maps from FV data to GLL and back, within an element, leave the FV data unchanged. Requirement R2 is the equivalent of R1 for mapping GLL data to FV and back. Because  $n_p \geq n_f$ , in general idempotence can hold only when the  $n_{p'}$ -basis exactly represents the  $n_p$ -basis field and  $n_{p'} = n_f$ . Requirement R2 assures the remap operator has order of accuracy (OOA)  $n_{p'} = n_f$  for  $C^{n_f-1}$  functions because a  $n_{p'}$ -basis-representable field is recovered exactly.

$A^{p \rightarrow f}$ , element locality, and the conditions in R2 uniquely specify  $A^{f \rightarrow p}$ . Assume the consequent, R2(d), of R2 holds:

$$A^{f \rightarrow p} \mathbf{f} = \mathbf{p}.$$

Then

$$\begin{aligned} A^{f \rightarrow p} A^{p \rightarrow f} \mathbf{p} &= \mathbf{p}, && \text{by R2(a),} \\ A^{f \rightarrow p} A^{p \rightarrow f} \mathcal{I}^{p' \rightarrow p} \mathbf{p}' &= \mathcal{I}^{p' \rightarrow p} \mathbf{p}', && \text{by R2(c),} \\ A^{f \rightarrow p} A^{p \rightarrow f} \mathcal{I}^{p' \rightarrow p} &= \mathcal{I}^{p' \rightarrow p}, && \text{because } \mathbf{p}' \text{ is general,} \\ A^{f \rightarrow p} &= \mathcal{I}^{p' \rightarrow p} (A^{p \rightarrow f} \mathcal{I}^{p' \rightarrow p})^{-1}. && \end{aligned} \quad (3)$$

The matrix inverted in the final line (3) is square because of R2(b). The formula (3) for  $A^{f \rightarrow p}$  satisfies R1:

$$A^{p \rightarrow f} A^{f \rightarrow p} \mathbf{f} = A^{p \rightarrow f} \mathcal{I}^{p' \rightarrow p} (A^{p \rightarrow f} \mathcal{I}^{p' \rightarrow p})^{-1} \mathbf{f} = \mathbf{f}.$$

In terms of the quantities in (2), we can write

$$A^{f \rightarrow p} = \mathcal{I}^{p' \rightarrow p} (\text{diag}(\mathbf{w}^f)^{-1} M^p \mathcal{I}^{p' \rightarrow p})^{-1}.$$

For convenience, let  $B^{f \rightarrow p} \equiv \text{diag}(\mathbf{J}^p)^{-1} A^{f \rightarrow p} \text{diag}(\mathbf{J}^f)$ .

### 2.2.3 FV to GLL: An alternative method for discontinuous fields

Mapping a discontinuous field from FV to GLL grids with large  $n_f$  produces oscillations. In future applications, we are interested in large  $n_f$  and  $n_p$ . Thus, we describe an additional linear remap operator  $\bar{A}^{f \rightarrow p}$  that addresses this case. In text we refer to the algorithm using the operator  $A^{f \rightarrow p}$  described in subsection 2.2.2 as the *high-order reconstruction* (HR) method and the algorithm using  $\bar{A}^{f \rightarrow p}$  as the *panel reconstruction* (PR) method. Outside of Section 2, only the HR method is used.

The PR method uses the HR method as a component. For each FV subcell in an element, the PR method reconstructs the field within the subcell by applying the HR method to a  $n_{f'} \times n_{f'}$  panel of nearby subcells, where  $n_{f'} \leq n_f$ . The panel's GLL basis has number of nodes  $n_p = n_{p'} = n_{f'}$ , which we summarize with the single parameter  $n_{f'}$ . A subcell's panel is determined by centering it on the subcell, then shifting it by the minimal amount necessary in each direction to place the panel within the bounds of the element. This first step produces a reconstruction that is generally discontinuous at subcell edges. The reconstruction has order of accuracy  $n_{f'}$  for  $C^{n_{f'}-1}$  functions because it is the result of applying the HR operator.

Then the PR method applies the  $L^2$  projection to project the subcell reconstructions onto the element's GLL  $n_p$ -basis. The  $L^2$  projection of a function  $y(\mathbf{r})$  over  $R$  is given by the solution of

$$\sum_j M_{ij}^{pp} p_j = \int_R b_i(\mathbf{r}) y(\mathbf{r}) \, d\mathbf{r}, \quad i = 1, \dots, n_p^2,$$

for  $\mathbf{p}$ , where  $M_{ij}^{pp} \equiv \int_R b_i(\mathbf{r}) b_j(\mathbf{r}) \, d\mathbf{r}$  is the GLL  $n_p$ -basis mass matrix. In the case of the PR method,  $y(\mathbf{r})$  is the function over the element resulting from the panel reconstructions of  $\mathbf{f}$ . The  $L^2$  projection preserves the order of accuracy of the panel reconstructions, giving an overall  $n_{f'}$ -order method.

Figure 2 shows an example of panels for the case of  $n_p = n_f = 5$ ,  $n_{f'} = 3$ . As in Figure 1, the blue solid line outlines the spectral element, the green filled circles mark GLL points, and the red dashed lines outline FV subcells. The black  $n_{f'} \times n_{f'}$  panel provides reconstructions for a  $2 \times 2$  grid of subcells in the southeast corner, marked with block dots. The yellow panel provides reconstructions for the two left-most subcells along the center row, marked with yellow plus signs. Finally, the cyan panel provides the reconstruction for the middle cell, marked with cyan  $\times$  signs.

If  $n_{f'} = n_f$ , then  $\bar{A}^{f \rightarrow p} = A^{f \rightarrow p}$ . First, every FV subcell has the same panel, and the panel covers the element. Second, the  $L^2$  projection from the panel's GLL  $n_{f'}$ -basis to the element's GLL  $n_p$ -basis is equivalent to interpolation by  $\mathcal{I}^{p' \rightarrow p}$  to the  $n_p$ -basis nodes since  $n_{f'} = n_{p'}$  and  $n_{f'} \leq n_p$ . Thus, in this case of  $n_{f'} = n_f$ , the HR and PR methods produce the same GLL solution.

The PR method satisfies neither R1 nor R2 if  $n_{f'} < n_f$ . First, R2 cannot hold because  $n_{f'} < n_f$ . Second, the  $L^2$  projection generally violates R1. However, conditions R1 and R2 still are used in deriving the panel reconstructions.

#### 2.2.4 Implementation

The  $A^{p \rightarrow f}$ ,  $A^{f \rightarrow p}$ , and  $\bar{A}^{f \rightarrow p}$  operators are linear and thus are computed once at model initialization; during time stepping, applications of these operators require small matrix-vector products with matrices of size  $n_f^2 \times n_p^2$  and  $n_p^2 \times n_f^2$ . Integrals are computed by Gaussian quadrature within each FV subcell. Since our application already has GLL nodes and weights of various degrees, we use GLL quadrature. For an integrand having polynomial degree  $d$ , the GLL basis with  $\lceil (d+3)/2 \rceil$  nodes is sufficient for exact integration.

### 2.3 Element-local property preservation

In the problem of *property preservation*, a property is a constraint that must hold nearly to machine precision, despite the approximations made in a discretization. We write “nearly” because in general the computation of the property has a condition number that, while small, is still larger than 1. For density and mixing ratio fields in the physgrid remap problem, the properties are respectively conservation of mass and nonviolation of the mixing ratio extrema in a discrete domain of dependence, or *shape preservation*. The second property also reestablishes tracer consistency if it does not already

hold; if mixing ratio  $q$  is constant in one basis, then the lower and upper extrema are the same and so  $q$  is mapped to a constant in the other.

For a density field  $\rho$ , the mass conservation constraint in an element is

$$\sum_i w_i^p J_i^p \rho_i^p = \sum_c w_c^f J_c^f \rho_c^f. \quad (4)$$

Both  $B^{p \rightarrow f}$  and  $B^{f \rightarrow p}$  conserve mass. The fact that  $B^{p \rightarrow f}$  conserves mass combined with requirement R1—and, in the case of the PR method, the additional fact that an  $L^2$  projection conserves mass—imply  $B^{f \rightarrow p}$  conserves mass.  $B^{p \rightarrow f}$  conserves mass as follows:

$$\begin{aligned} \sum_c w_c^f J_c^f \rho_c^f &= \sum_c w_c^f J_c^f (B^{p \rightarrow f} \boldsymbol{\rho}^p)_c = \sum_c w_c^f J_c^f (J_c^f)^{-1} (w_c^f)^{-1} (M^p \text{diag}(\boldsymbol{J}^p) \boldsymbol{\rho}^p)_c \\ &= \sum_c (M^p \text{diag}(\boldsymbol{J}^p) \boldsymbol{\rho}^p)_c = \sum_c \sum_i M_{ci}^p J_i^p \rho_i^p = \sum_c \sum_i \int_{R_c^f} b_i(\mathbf{r}) d\mathbf{r} J_i^p \rho_i^p \\ &= \sum_i J_i^p \rho_i^p \sum_c \int_{R_c^f} b_i(\mathbf{r}) d\mathbf{r} = \sum_i w_i^p J_i^p \rho_i^p. \end{aligned}$$

Because the  $B$  operators are linear and implement high-order remap, they cannot preserve shape in general. To preserve shape, we apply an element-local limiter after  $B$ . In each element, this limiter perturbs input mixing ratio  $\mathbf{q}$  to produce output  $\mathbf{q}'$  such that mass is conserved,  $\sum_i w_i J_i \rho_i q_i = \sum_i w_i J_i \rho_i q'_i$ ;  $q'$  is in bounds; and the difference  $\text{diag}(\boldsymbol{\rho}) \mathbf{q}' - \text{diag}(\boldsymbol{\rho}) \mathbf{q}$  is 1-norm-minimal over all modifications that obey these constraints. It can be applied to a field in an element in either basis. There are a number of limiters that provide such outputs  $\mathbf{q}'$ ; we use CLIPANDASSURED SUM, Algorithm 3.1 in [2].

In the physgrid remap problems, the mass conservation and shape preservation constraints always form a non-empty set. The shape preservation constraint set can be written

$$\min_i q_i^s \leq q_j^t \leq \max_i q_i^s,$$

where  $s$  denotes a quantity on the source grid and  $t$  on the target grid.  $j$  ranges over the target degrees of freedom (DOF) within the element, and  $i$  ranges over the source DOF in the element as well as possibly neighbors. A constant mixing ratio over the element satisfies these constraints; thus, the constraint set is not empty.

## 2.4 Global grids

Now we discuss the steps to remap quantities from one global grid to the other.

In HOMME, the total mass in a model layer is given by the *pseudodensity* times the vertical coordinate differential, where pseudodensity is the derivative of hydrostatic pressure with respect to the vertical coordinate; see [24] for details, in which the pseudodensity is written  $\partial p / \partial \eta$ . For purposes of this discussion, which is independent of the details of HOMME, it is sufficient to refer to a general density  $\rho$  whose integral over a volume yields a mass. We refer to scalars  $s$  and corresponding scalar densities  $s\rho$ , as well as scalar tendency  $\Delta s$  and scalar density tendency  $\Delta(s\rho)$ .

Remapping a scalar from GLL to FV bases is purely an element-local operation. If the limiter will be applied, the bounds are the extremal scalar values of the GLL nodal values. First, the scalar density field is remapped. This same remap operator is applied to the total mass density; thus, if the scalar is a tracer mixing ratio, then it is mass-tracer consistent already. Second, the limiter is applied to the scalar to limit new extrema. Each of these two steps is local to the element.

In the FV-to-GLL direction, the temperature  $T^f$  and velocity  $u^f, v^f$  data are tendencies. Tracer  $q^f$  is typically a full state, but to avoid dissipating tracers by repeatedly remapping state, just the tracer tendency is remapped. The limiter is applied to tracers but not  $T^f, u^f, v^f$ ; in what follows, the limiter steps are simply omitted for these. For concreteness, the following algorithm description refers to tracers in particular.

The algorithm proceeds as follows. To solve the correct property-preservation problem, limiter bounds must be for the full GLL tracer state. If the tracer density tendency is zero, the limiter should not alter the state. Thus, the bounds must include the extremal nodal values of  $q^p$  prior to state update. First, within each element, the minimum and maximum (subsequently, *extremal*) FV mixing ratio state values are computed and stored. Second, in each element, the tracer density tendency is remapped with the linear, element-local, remap operator. Third, to permit OOA 2, a halo exchange shares the FV mixing ratio state extremal values among directly neighboring elements. After the exchange, the bounds for an element's mixing ratio state are the extremal values from the element and its neighbors. These are then augmented by the state's previous GLL extremal values in the element. Fourth, the limiter is applied to the mixing ratio values. Fifth, the DSS is applied. After the DSS, mixing ratio extrema within an element are bounded by the source data within a 2-halo of the element rather than the original 1-halo. The DSS is globally tracer mass conserving, but an element's tracer mass after the DSS is altered in general.

Returning to conditions R1 and R2 in subsection 2.2.2, neither condition holds in general throughout the full algorithm. Both property preservation and the DSS can violate each condition. These conditions were used just to derive linear, element-local, remap operators.

Figure 3 illustrates the steps of the FV-to-GLL remap procedure for a tracer. The left column uses the HR operator and the right uses the PR operator. The black stair-step lines show the FV state. Vertical gray lines mark FV subcell boundaries, and black vertical dashed lines mark spectral element boundaries. In total there are four spectral elements over the domain, with periodic boundary conditions on the domain edges. The parameters are  $n_p = 8, n_f = 5$ . In all figure parameter labels, we omit '=' and ',', trading this space for a larger font size; e.g., " $n_p = 8, n_f = 5$ " is written " $n_p\ 8\ n_f\ 5$ ". The FV data are  $C^1$  continuous on the right side of the domain and discontinuous on the left. For simplicity, the previous step's state is the constant blue dotted line, labeled as "(0) previous step", and the total mass density is uniformly one (not shown). The difference between the FV stair-step curve and this blue dotted line is then the tendency to remap.

In the top row, the linear, element-local, remap operator is applied to produce the dotted green curve labeled "(3) DGLL  $n_p\ 8$ ". Here, "D" in "DGLL" denotes *discontinuous*; at this stage, the GLL field is discontinuous across elements. Large green filled circles mark GLL nodes. Note that element boundary nodes are two-valued. The red curves labeled with a "(2)" show the intermediate quantity in each of the HR (left) and PR (right) methods. In the HR case, the intermediate quantity is the DGLL  $n_{p'} = n_f = 5$  reconstruction. In the PR case, the intermediate quantity is the function formed by panel reconstructions with  $n_{f'} = 3$ .

Moving now to the second row, CLIPANDASSURED SUM (CAAS) is applied to the green dotted curve to produce the green dashed curve labeled "(4) DGLL". The right column uses two sets of arrows to show how the bounds are obtained. The green upward arrows show the elements whose FV state data are used to provide an initial set of bounds for the element pointed to by the downward green arrow. The blue downward arrow shows that this initial set of bounds is augmented by the extremal values in the previous GLL state of that element.

Finally, the DSS is applied to the green dashed curve to produce the solid blue curve in the bottom row labeled "(5) CGLL". Here, the "C" denotes *continuous*; the GLL field is now continuous across elements. For reference, the GLL nodes are marked with red dots. Visually, the HR and PR methods reconstruct the smooth part of the tendency similarly, but the PR method produces lower-magnitude oscillations than the HR method across the discontinuity.

## 2.5 Remapped quantities

Until now, this section has described remap algorithms general to any discretization involving spectral elements with embedded finite volume subcells. This subsection describes the quantities remapped in the E3SM atmosphere model.

Tracer mixing ratios are remapped using the methods described in subsections 2.2, 2.3, and 2.4.

Potential temperature density  $\theta\rho$  is linearly remapped, in the GLL-to-FV direction as a full state and in the FV-to-GLL direction as a tendency, but no limiter is applied, consistent with its treatment in the dynamical core. Experiments show low sensitivity to the particular temperature variable type that is remapped. We choose potential temperature because it is conserved under dry adiabatic motion; that is, the integral of  $\theta\rho$  in a Lagrangian parcel undergoing dry adiabatic motion is conserved. However, virtual potential temperature or just temperature could be used.

Velocities  $u$ ,  $v$  are respectively the local-east and local-north velocity components. At and near the poles, these vary quickly in space and so will remap with more error than away from the poles. Instead of remapping these, we map the velocity vector to a reference-element coordinate system, remap each component in this coordinate system between grids—thus conserving the component’s average value in this coordinate system—then map the result back to the local east-north system. Examples of element-local coordinate systems are contravariant, covariant, and 3D Cartesian. In our implementation, we use the contravariant coordinate system. A limiter is not applied to velocity, consistent with its treatment in the dynamical core.

There is no vertical velocity tendency, but the physics parameterizations need access to the vertical velocity  $\omega$ . This quantity’s state is remapped as a scalar from GLL to FV grids, conserving  $\omega$ ’s average value over a subcell, and no limiter is applied.

Topography data, geopotential  $\phi_s^p$  and  $\phi_s^f$ , are preprocessed and stored in a topography file.  $\phi_s^p$  are the traditional GLL  $\phi_s$  values; these are used in the dynamical core.  $\phi_s^f$  are obtained by remapping from GLL to FV and then applying the limiter; these are used in the physics parameterizations. Remap prior to application of the limiter preserves average elevation within a subcell; after, within an element. Elevation variance and other derived data used in physics parameterizations are computed offline using  $\phi_s^f$  and stored in the same topography file. See [15] for details on these derived topography data.

One detail specific to HOMME is the treatment of hydrostatic surface pressure  $p_s$ , which is needed in the physics parameterizations, and the remap of pseudodensity, which is needed for remap of other quantities. A full description of these quantities can be found in [24]. The three key points for our purposes are as follows. First, the integral of pseudodensity over a column gives the hydrostatic surface pressure  $p_s$ . Second, given surface pressure, HOMME provides the vertical coordinate data necessary to compute pseudodensity over the column. Third, from the pseudodensity one can compute hydrostatic pressure at vertical column interfaces and midpoints. None of these three points involves remap. In both remap directions, we must simply remap  $p_s^p$  to obtain  $p_s^f$ , then use HOMME’s existing vertical coordinate treatment to derive pseudodensity on the FV grid.

## 2.6 Numerical verification

We verify our remap operators using a convergence study on a sequence of quasiuniform cubed-sphere grids. A tracer field is created on the GLL grid, remapped to the FV grid, and then remapped back to the GLL grid. Then the relative error between original and remapped fields is measured. We repeat this procedure on a sequence of increasingly fine grids. The tracer fields are described in [14]: Gaussian hills, cosine bells, and slotted cylinders. These differ in smoothness:  $C^\infty$ ,  $C^1$ , and discontinuous, respectively. The grid sequence is expressed in terms of  $n_e$ : 5 to 320, increasing at each step by a factor of 2. We test a number of configurations. When we use the PR method, we set

$n_{f'} = \min(n_f, 3)$ . Because the HR and PR methods are identical when  $n_{f'} = n_f$ , we label those cases as “HR/PR” to indicate that either method would produce the result.

Figure 4 shows results. The legend describes the line pattern, color, and marker for each algorithm. Triangles provide reference order of accuracy (OOA) slopes, with OOA in text in the triangle.

The green dotted line with  $\times$  markers shows that on a sufficiently smooth field and with no limiter, the OOA of the HR method is  $n_f$  as a result of condition R2. The blue solid line corresponds to our intended practical configuration:  $n_f = 2$  with limiter on for tracer fields. It achieves OOA 2 on a  $C^1$  field. The red solid line shows that for a field having a finite number of extrema, the limiter permits OOA 3. This is because the limiter is triggered in only  $O(1)$  elements in this case. The green solid line is for  $n_p = n_f = 4$ , HR method, with limiter. It is jagged, but still monotonically decreasing, in the case of Gaussian hills because the limiter is triggered at just the two isolated maxima and is sensitive to the details of these maxima with respect to the grid. But the accuracy is always between the curves for  $n_p = 5$ ,  $n_f = 4$ , HR method, with no limiter (dotted green) and  $n_p = 4$ ,  $n_f = 3$ , HR/PR, with limiter (solid red), as we expect. For continuous but less smooth functions for which the limiters are triggered in many elements, the solid green line is straight. For the discontinuous slotted cylinders test field, OOA is sublinear in the  $l_2$  norm, as we expect. In the  $l_\infty$  norm, error is roughly constant with refinement because of roughly constant-magnitude under- and overshoots at the discontinuity. We do not show curves for the HR method with  $n_p = n_f$  and no limiter; they are roughly at machine precision because of requirements R1 and R2.

For the cases in which the PR method produces different results than the HR method—lines with unfilled markers—results are again as expected. For the Gaussian hills field,  $n_p = n_f = 4$ , PR, with limiter (dashed green line with unfilled circles), the OOA is 3 because  $n_{f'} = 3$ . For the case with  $n_p = 8$  (dashed black line with unfilled triangles), the explanation above regarding jaggedness applies. We omit curves using the PR method without limiter because the OOA is limited to 3, and the dashed green line with unfilled circles already demonstrates OOA 3.

Figure 5 shows contour plots for a subset of the cases in Figure 4, all for  $n_e = 20$ , which is the typical coarser resolution of the two resolutions used in studies involving these test fields. All results shown are with CAAS applied. Each row of the figure corresponds to a field, labeled in the left column of the figure. To maximize detail in the figure, each plot zooms in on one of the two prescribed shapes within a field. The left column shows the original field. Each subsequent column shows the results of applying the algorithm described at the top of the column. For the Gaussian hills and cosine bells fields, the logarithm, base 10, of the magnitude of the difference between final GLL field and original is shown, since showing the final field itself would provide little information. There are two different color scales used for these two rows: one for the first column, and the other for the remaining columns. For the slotted cylinders field, the final field is sufficiently different that it is informative to show it directly, and the colorbar applies to all columns. Some observations of interest are as follows. For the Gaussian hills and cosine bells fields, the errors for the case  $n_p = n_f = 4$ , HR method, center on the manifolds where the CAAS limiter is active: the Gaussian hills maximum, a point, and the cosine bells border, a circle. The PR method has larger errors for this  $n_p = n_f = 4$  configuration, as expected. But the PR method is motivated by the slotted cylinders result for  $n_f \geq 4$ . With the HR method, we see oscillations in the elements that overlap the border of the slotted cylinder in the case  $n_f = 4$ . With the PR method, these are substantially reduced. The case  $n_p = 8$ ,  $n_f = 6$ , PR method, shows that the PR method with  $n_{f'} = 3$  suppresses oscillations resulting from discontinuous fields for large  $n_f$ .

### 3 Experiment Setup

#### 3.1 Model Simulations

E3SMv1 was originally forked from the NCAR CESM version 1 [11] and has continued to evolve with a focus on targeting next-generation exascale computers. Some notable differences between the atmosphere component in E3SM and CESM are the use of spectral element dynamics on a cubed-sphere grid [20, 24], more vertical levels (72 vs. 32), and a higher model top (approx. 60 km vs. 40 km). The various physics schemes are broadly similar, but many details differ in important ways [27, 18]. The land, ocean, and sea ice components also differ significantly from CESM. The simulations presented here use active atmosphere and land components with prescribed sea surface temperature (SST) and sea ice.

To analyze the impact of the physgrid on E3SM, four simulations were conducted with  $30 \times 30$  elements per cube face (ne30). The physics grids for these experiments are ne30np4, ne30pg2, ne30pg3, and ne30pg4. We also conducted np4 and pg2 simulations that include a  $4 \times$  mesh refinement over the continental US similar to [22]. The high-order reconstruction (HR) method for FV-to-GLL remap was used in all simulations. All simulations were run on Cori-KNL at NERSC using 2 OpenMP threads per MPI rank. The ne30 simulations used 85 nodes (5400 MPI ranks) and the regionally refined cases used 150 nodes (9600 MPI ranks). An additional experiment was run with ne45pg2, which has almost the same number of physics columns as the ne30np4 case, to explore whether the effective resolution of the model is affected by the physgrid. All simulations were run for a total of 5 years. The use of 5-year simulations is a common practice in model development and is a trade-off between computational cost and signal-to-noise ratio, but there can still be low-frequency variability that can influence the results. Additional 5-day aqua-planet experiments with file output disabled were run to help assess the performance implications of the physgrid in section 4.4.

Since our primary goal is to compare the impact of changing the physics grid we want to use the same grid for the surface component across all simulations. The land model grid is often specified to match the atmosphere grid, but for our purposes the land model is run on an equiangular  $0.5^\circ$  grid for all simulations. The ocean and ice components of E3SM [17] use an unstructured mesh based on Voronoi tessellations [19], and all configurations share the same grid for the ocean and sea ice components even though these components only provide prescribed conditions in the simulations here.

The model input data for quantities like solar forcing and land surface types is derived to be representative of climatological conditions around 2010, similar to [3]. The prescribed annual cycle of SST, sea ice concentrations, and aerosol emissions are derived from a 10-year climatology over 2005-2014. Greenhouse gas concentrations are taken from the 2010 annual mean. The atmospheric initial condition is stored on the dynamics grid and is therefore the same for all simulations with the same underlying dynamics grid. The atmospheric initial condition was taken from year 2010 of an AMIP-style run.

Figure 6 illustrates the differences between the primary grids considered in this study using topography data focused on the continental United States for the ne30np4, ne30pg2, ne30pg3, and ne30pg4 grids. Each point of the np4 grid represents a nodal point of the spectral element grid that does not correspond to a specific area on the sphere, thus data cannot be visualized directly like cell averaged data from an FV grid. To create a finite volume approximation of the np4 data we use an iterative process to define a quadrilateral centered around each np4 point. The area of this quadrilateral matches the GLL quadrature weight of the node so that area-weighted global averages can be calculated. All grids capture the same general topographical features, but the pg4 grid naturally captures the most detail. The regionally refined grid is highlighted in Figure 7 for the pg2 cases. The mesh refinement is only considered at the element level, so both np4 and pg2 regionally refined cases start with the same refinement pattern to define the sub-element physics grid.

### 3.2 Model Validation

Our primary goal is to test that the physgrid in E3SM does not degrade the simulation relative to the configuration with physics columns on the GLL nodes, and for that purpose we can focus on comparing the simulations with each other. However, it is also useful to put our results in the context of previously published model intercomparisons, so we additionally compare our simulations to E3SMv1 AMIP experiments from G19 and data from the Coupled Model Intercomparison Project Phase 5 (CMIP5) [23]. To make this comparison concise we compute spatial RMS errors relative to observations for annual and seasonal averages with the PCMDI Metrics Package [6]. The observation data for these calculations is provided by the developers of the metrics package.

## 4 Results

### 4.1 Climatology

To examine the simulation results we start with a simple comparison of the 5 year mean precipitation, liquid water path, total column water vapor, and 500 mb zonal wind in Figure 8 shown as differences relative to the ne30np4 case. Similar analysis of the regionally refined cases is also shown in Figure 9. The data in all cases was regridded to a common 1 degree grid to facilitate a direct comparison. Statistical significance was determined by testing whether the differences of means were significantly different than zero with a two-tailed Student's *t*-test at a 95% confidence level, which is indicated with stippling in Figure 8. The degrees of freedom for significance testing at each location were determined to be 5 by reasoning that each year of data makes up an independent and identically distributed sample of the true annual mean climatology.

The climatological patterns of the physgrid cases are largely similar to their np4 counterparts. Some statistically significant regional differences are evident, but the spatial patterns of these differences are not consistent across physgrid cases, so we conclude that they do not indicate a fundamental change in model behavior driven by the change in the physics grid. Furthermore, small systematic differences are not completely unexpected from the change in the physics grid. Some of the small scale differences, such as the dry bias near the maritime continent in pg2 and pg3 cases, may be related to how coastlines and topography may be slightly different on the physgrid. The differences are also likely due in part to low frequency variability that still occurs in simulations with a prescribed annual cycle of SSTs. Analysis of other quantities produced similar results (not shown).

For a more comprehensive comparison of the physgrid results we recreate the root-mean-square error analysis from G19 (see their Fig. 9) in Figure 10, which compares the simulations presented here with 45 CMIP5 models (boxes and whiskers) and the AMIP simulation from G19 (red markers). The RMSE values for all simulations are within or below the envelope of CMIP5 results, and the physgrid simulations produce similar results to the baseline np4 case. Smaller RMSE values in these simulations relative to the CMIP5 data are not surprising due to the lack of systematic biases in the SST. Similarly, the slight deviations between our simulations and the G19 data are also not surprising because the G19 simulation used transient SST forcing that includes inter-annual variability, whereas the simulations here used a prescribed annual cycle of SST. Overall the RMSE analysis gives further confidence that the physgrid implementation does not qualitatively alter the solution of E3SM.

### 4.2 GLL Grid Noise

An interesting result of [9] is that noise in the vertical pressure velocity field (i.e. omega) around mountainous terrain that is present in both np4 and pg3 configurations was eliminated on the physics

grid of the the pg2 case. To investigate this in E3SM Figure 11 shows the time mean omega at 500 mb on both the physics and dynamics grids, referred to as OMEGA and DYN\_OMEGA, respectively.

Figure 11a-d shows that the noise over mountainous regions is eliminated in the physics grid data of the pg2 case, similar to the results of [9]. Interestingly, the noise signal is still present on the underlying dynamics grid, although slightly reduced in magnitude for the pg2 case (Fig. 11e-h). Thus, the lack of noise on the ne30pg2 physics grid is mainly a result of smoothing by the coarseness of the physics grid relative to the dynamics grid. While it is good to be aware of this underlying noise when using the pg2 grid, we do not see this as a major issue since most model analysis will only consider data on the physics grid, although we recognize that it could still be having an undesirable, albeit small, influence on the model behavior.

### 4.3 Effective Resolution

Estimating the effective resolution of a spectral model is not a straightforward task [13, 21], but here we are only interested in any relative change of the effective resolution by the physgrid that might affect our interpretation of the model results. The mapping from dynamics to the pg2 grid acts as a low-pass filter on the state, but it is not immediately obvious how a coarser physics resolution will feed back onto the dynamics.

To investigate this question we first calculate spherical harmonic power spectra of physics temperature tendencies on the dynamics grid of the ne30 simulations using a model level near 500 mb (Fig. 12). Using the physics tendencies on the dynamics grid provides a direct comparison to the baseline ne30np4 simulation data because the same map from model grid to the grid used to compute spherical harmonics is used for all cases. The spectra were calculated from daily snapshots after remapping to a 721 by 1440 Gaussian grid to facilitate a spherical harmonic transform using a high order, conservative method [26]. The spectra of physics tendencies in Figure 12 indicate a lower effective resolution with pg2 as expected (dashed red line), which shows a slightly faster roll off of power with increasing wave number. The pg3 and pg4 cases (blue and green dashed line) more closely resemble the np4 baseline (black line).

While the spectra of physics tendencies directly show the effect of the mapping, we can also use the power spectra of kinetic energy to examine how the change in physics tendencies affects the behavior of the dynamics and how this is ultimately represented on the physics grid that is typically used for analysis. Figure 13 shows spherical harmonic power spectra of kinetic energy using physics grid data for the ne30np4, ne30pg2, and ne45pg2 cases and the dynamics grid data for the ne30pg2 case. The data were similarly remapped to a 721 by 1440 Gaussian grid using a high order, conservative method for GLL data and a bilinear method for FV physgrid data.

The effective resolution can be estimated from the roll off of the kinetic energy spectra at higher wave numbers. The ne30pg2 case begins to diverge from the ne30np4 case around wave number 30–40, suggesting a slightly lower effective resolution that should be expected from the lower degrees of freedom on the physgrid. However, the dashed red curve shows the spectra of dynamics grid data from the ne30pg2 case and lies directly on top of the ne30np4 spectra. So we conclude that the use of the physgrid does not degrade the effective resolution of the dynamics even though the representation of the model state on the physgrid is smoother.

The ne45pg2 case provides an insightful comparison because it has almost the same number of physics columns as the ne30np4 case (48600 vs. 48602). We might naively expect a similar effective resolution between the two, but the spectra in Figure 13 show that this is not the case. The ne45pg2 data show that the effective resolution is still higher than ne30np4 due to the higher resolution of the underlying ne45np4 dynamics grid.

#### 4.4 Model Performance

To summarize the performance impact of the physgrid Figure 14a shows individual estimates of overall model throughput (small gray dots) and the average throughput over the whole simulation (large black dots) in simulated years per wall-clock day (sypd). The overall throughput estimates include the effects of interprocess communication and file output. If we only consider timing estimates of the dynamics calculations and convert these data into throughput values the results are roughly the same across cases that share the same underlying dynamics grid (Fig 14b). For the same throughput conversion of physics timing estimates we see much more variation between cases due to the changing number of physics columns (Fig 14c). The pg2 case has significantly higher throughput due to having  $4/9$  as many physics columns, whereas the pg4 case has lower throughput due to having  $16/9$  as many physics columns. The pg3 case has almost the same number of physics columns as np4, which explains why the throughput is similar.

The spread of the individual dynamics throughput estimates is primarily caused by the interprocess communication required for the direct stiffness summation (DSS) that communicates information between elements. The individual physics throughput estimates are much more consistent due to the lack of interprocess communication. In spite of this fact, the physics calculations still do not line up with the idealized scaling we might expect from simply changing the number of columns (not shown).

Figure 15 shows the throughput of 5-day aquaplanet simulations with file output disabled using 85, 43, and 22 Cori-KNL nodes. Colored lines show the ideal throughput based on scaling the results of the ne30np4 grid by the change in the physics columns for each physgrid case. Figure 15 illustrates that the departure from the ideal scaling is a result of using more nodes for higher throughput. A smaller number of nodes shows much better agreement with the ideal scaling, which highlights a loss of efficiency when running in a higher throughput regime. This is consistent with the results of [1] who showed that Cori-KNL efficiency can vary with the amount of workload per node.

### 5 Conclusions

This study details the implementation and impact of a quasi-equal-area finite volume physics grid for use in E3SM alongside the spectral element dynamics grid. The mapping between the dynamics and physics grids is constrained to be local to each spectral element to minimize inter-process communication and permit regional mesh refinement. The motivation to provide an alternative to the more common co-located grids approach was both to avoid the effects of grid imprinting and to improve performance by reducing the number of physics columns for a given dynamics grid.

While the new physics grid can be specified to have a finer average spacing than the dynamics grid with  $4 \times 4$  finite volume cells per element (pg4), the model solution is shown to be qualitatively insensitive to this change. This corroborates the results of [9] who found that the resolved scale of the model is primarily determined by the effective resolution of the dynamics. Thus, a physics grid with  $2 \times 2$  finite volume cells per element (pg2), which is coarser than the underlying dynamics grid, is preferable to optimize performance. Simulations show that the physics throughput can be increased by approximately 60% to 120%, depending on configuration of computational resources, using the pg2 grid over the co-located np4 grid.

Analysis of the simulated climate does not indicate any large systematic change in the simulated climate due to the new physics grid, although noise over steep topography is notably reduced with pg2. In general, grid noise and imprinting is not easily detectable away from steep topography when using traditional physics parameterizations, but the noise becomes significantly amplified when using the Multi-scale Modelling Framework (MMF) configuration of E3SM (E3SM-MMF) [8], in which a cloud resolving model is embedded in each physics column to explicitly represent convective scale processes.

The use of the pg2 grid with E3SM-MMF helps remedy the grid imprinting problem in addition to taming the substantial increase in physics cost associated with the MMF.

**Acknowledgements.** This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration, and by the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and 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-NA-0003525. 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 research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. The E3SM project, code, simulation configurations, model output, and tools to work with the output are described at the website (<https://e3sm.org>). Instructions on how to get started running E3SM are available at the website (<https://e3sm.org/model/running-e3sm/e3sm-quick-start>). All code for E3SM may be accessed on the GitHub repository (<https://github.com/E3SM-Project/E3SM>). The code for the branch of E3SM used in this study and the raw output data is archived at the DOE's National Energy Research Scientific Computing Center (NERSC). The code used to generate the verification data in Section 2 is at <https://github.com/E3SM-Project/COMPOSE/releases/tag/v1.1> in the `methods` directory. Additional code, notes, and a subset of data that can be used to reproduce our results is archived at [https://dabdceba-6d04-11e5-ba46-22000b92c6ec.e.globus.org/publications/Hannah\\_JAMES\\_2021.tar.gz](https://dabdceba-6d04-11e5-ba46-22000b92c6ec.e.globus.org/publications/Hannah_JAMES_2021.tar.gz). SAND2021-7186 J.

## References

- [1] Luca Bertagna, Michael Deakin, Oksana Guba, Daniel Sunderland, Andrew M Bradley, Irina K Tezaur, Mark A Taylor, and Andrew G Salinger. HOMMEXX 1.0: a performance-portable atmospheric dynamical core for the Energy Exascale Earth System Model. *Geoscientific Model Development*, 12(4):1423–1441, 2019.
- [2] Andrew M. Bradley, Peter A. Bosler, Oksana Guba, Mark A. Taylor, and Gregory A. Barnett. Communication-efficient property preservation in tracer transport. *SIAM Journal on Scientific Computing*, 41(3):C161–C196, 5 2019.
- [3] Peter M. Caldwell, Azamat Mametjanov, Qi Tang, Luke P. Van Roekel, Jean Christophe Golaz, Wuyin Lin, David C. Bader, Noel D. Keen, Yan Feng, Robert Jacob, Mathew E. Maltrud, Andrew F. Roberts, Mark A. Taylor, Milena Veneziani, Hailong Wang, Jonathan D. Wolfe, Karthik Balaguru, Philip Cameron-Smith, Lu Dong, Stephen A. Klein, L. Ruby Leung, Hong Yi Li, Qing Li, Xiaohong Liu, Richard B. Neale, Marielle Pinheiro, Yun Qian, Paul A. Ullrich, Shaocheng Xie, Yang Yang, Yuying Zhang, Kai Zhang, and Tian Zhou. The DOE E3SM Coupled Model Version 1: Description and Results at High Resolution. *Journal of Advances in Modeling Earth Systems*, 11(12):4095–4146, 12 2019.
- [4] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. Zang. *Spectral Methods: Evolution to Complex Geometries and Applications to Fluid Dynamics*. Springer, 1rst edition edition, 2007.

- [5] John M. Dennis, Jim Edwards, Katherine J. Evans, Oksana Guba, Peter H. Lauritzen, Arthur A. Mirin, Amik St-Cyr, Mark A. Taylor, and Patrick H. Worley. CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *The International Journal of High Performance Computing Applications*, 26(1):74–89, 2 2012.
- [6] Peter J. Gleckler, Charles Doutriaux, Paul J. Durack, Karl E. Taylor, Yuying Zhang, Dean N. Williams, Erik Mason, and Jrme Servonnat. A more powerful reality test for climate models. *Eos (United States)*, 97(12):20–24, 6 2016.
- [7] JeanChristophe Golaz, Peter M. Caldwell, Luke P. Van Roekel, Mark R. Petersen, Qi Tang, Jonathan D. Wolfe, Guta Abeshu, Valentine Anantharaj, Xylar S. AsayDavis, David C. Bader, Sterling A. Baldwin, Gautam Bisht, Peter A. Bogenschutz, Marcia Branstetter, Michael A. Brunke, Steven R. Brus, Susannah M. Burrows, Philip J. CameronSmith, Aaron S. Donahue, Michael Deakin, Richard C. Easter, Katherine J. Evans, Yan Feng, Mark Flanner, James G. Foucar, Jeremy G. Fyke, Brian M. Griffin, Ccile Hannay, Bryce E. Harrop, Elizabeth C. Hunke, Robert L. Jacob, Douglas W. Jacobsen, Nicole Jeffery, Philip W. Jones, Noel D. Keen, Stephen A. Klein, Vincent E. Larson, L. Ruby Leung, HongYi Li, Wuyin Lin, William H. Lipscomb, PoLun Ma, Salil Mahajan, Mathew E. Maltrud, Azamat Mametjanov, Julie L. McClean, Renata B. McCoy, Richard B. Neale, Stephen F. Price, Yun Qian, Philip J. Rasch, J.E. Jack Reeves Eyre, William J. Riley, Todd D. Ringler, Andrew F. Roberts, Erika L. Roesler, Andrew G. Salinger, Ze-shawn Shaheen, Xiaoying Shi, Balwinder Singh, Jinyun Tang, Mark A. Taylor, Peter E. Thornton, Adrian K. Turner, Milena Veneziani, Hui Wan, Hailong Wang, Shanlin Wang, Dean N. Williams, Philip J. Wolfram, Patrick H. Worley, Shaocheng Xie, Yang Yang, JinHo Yoon, Mark D. Zelinka, Charles S. Zender, Xubin Zeng, Chengzhu Zhang, Kai Zhang, Yuying Zhang, Xue Zheng, Tian Zhou, and Qing Zhu. The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, page 2018MS001603, 3 2019.
- [8] W. M. Hannah, C. R. Jones, B. R. Hillman, M. R. Norman, D. C. Bader, M. A. Taylor, L. R. Leung, M. S. Pritchard, M. D. Branson, G. Lin, K. G. Pressel, and J. M. Lee. Initial Results From the SuperParameterized E3SM. *Journal of Advances in Modeling Earth Systems*, 12(1), 1 2020.
- [9] Adam R. Herrington, Peter H. Lauritzen, Kevin A. Reed, Steve Goldhaber, and Brian E. Eaton. Exploring a lower resolution physics grid in CAMSECSLAM. *Journal of Advances in Modeling Earth Systems*, page 2019MS001684, 5 2019.
- [10] Adam R. Herrington, Peter H. Lauritzen, Mark A. Taylor, Steve Goldhaber, Brian E. Eaton, Julio T. Bacmeister, Kevin A. Reed, and Paul A. Ullrich. PhysicsDynamics Coupling with Element-Based High-Order Galerkin Methods: Quasi-Equal-Area Physics Grid. *Monthly Weather Review*, 147(1):69–84, 1 2019.
- [11] James W. Hurrell, M. M. Holland, P. R. Gent, S. Ghan, Jennifer E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, S. Marshall, James W. Hurrell, M. M. Holland, P. R. Gent, S. Ghan, Jennifer E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, and S. Marshall. The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360, 9 2013.

- [12] J. Lander, B. J. Hoskins, J. Lander, and B. J. Hoskins. Believable Scales and Parameterizations in a Spectral Transform Model. [http://dx.doi.org/10.1175/1520-0493\(1997\)125;2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1997)125;2.0.CO;2), 2 1997.
- [13] Ren Laprise. The resolution of global spectral models. *Bulletin of the American Meteorological Society*, 73(9):1453–1455, 9 1992.
- [14] P. H. Lauritzen, W. C. Skamarock, M. J. Prather, and M. A. Taylor. A standard test case suite for two-dimensional linear transport on the sphere. *Geoscientific Model Development*, 5(3):887–901, 6 2012.
- [15] Peter H Lauritzen, Julio T Bacmeister, Patrick F Callaghan, and Mark A Taylor. NCAR global model topography generation software for unstructured grids. *Geoscientific Model Development Discussions*, 8(6):4623–4651, 2015.
- [16] Peter Hjort Lauritzen, Mark A. Taylor, James Overfelt, Paul A. Ullrich, Ramachandran D. Nair, Steve Goldhaber, and Rory Kelly. CAM-SE-CSLAM: Consistent coupling of a conservative semi-lagrangian finite-volume method with spectral element dynamics. *Monthly Weather Review*, 145(3):833–855, 3 2017.
- [17] Mark R. Petersen, Xylar S. Asay-Davis, Anne S. Berres, Qingshan Chen, Nils Feige, Matthew J. Hoffman, Douglas W. Jacobsen, Philip W. Jones, Mathew E. Maltrud, Stephen F. Price, Todd D. Ringler, Gregory J. Streletz, Adrian K. Turner, Luke P. Van Roekel, Milena Veneziani, Jonathan D. Wolfe, Phillip J. Wolfram, and Jonathan L. Woodring. An Evaluation of the Ocean and Sea Ice Climate of E3SM Using MPAS and Interannual CORE-II Forcing. *Journal of Advances in Modeling Earth Systems*, 11(5):1438–1458, 5 2019.
- [18] P.J. Rasch, S. Xie, P.L. Ma, W. Lin, H. Wang, Q. Tang, S.M. Burrows, P. Caldwell, K. Zhang, R.C. Easter, P. CameronSmith, B. Singh, H. Wan, J.C. Golaz, B.E. Harrop, E. Roesler, J. Bacmeister, V.E. Larson, K.J. Evans, Y. Qian, M. Taylor, L.R. Leung, Y. Zhang, L. Brent, M. Branstetter, C. Hannay, S. Mahajan, A. Mametjanov, R. Neale, J.H. Richter, J.H. Yoon, C.S. Zender, D. Bader, M. Flanner, J.G. Foucar, R. Jacob, N. Keen, S.A. Klein, X. Liu, A.G. Salinger, M. Shrivastava, and Y. Yang. An Overview of the Atmospheric Component of the Energy Exascale Earth System Model. *Journal of Advances in Modeling Earth Systems*, page 2019MS001629, 7 2019.
- [19] Todd Ringler, Mark Petersen, Robert L. Higdon, Doug Jacobsen, Philip W. Jones, and Mathew Maltrud. A multi-resolution approach to global ocean modeling. *Ocean Modelling*, 69:211–232, 9 2013.
- [20] C. Ronchi, R. Iacono, and P.S. Paolucci. The Cubed Sphere: A New Method for the Solution of Partial Differential Equations in Spherical Geometry. *Journal of Computational Physics*, 124(1):93–114, 3 1996.
- [21] William C. Skamarock. Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132(12):3019–3032, 12 2004.
- [22] Qi Tang, Stephen A. Klein, Shaocheng Xie, Wuyin Lin, Jean Christophe Golaz, Erika L. Roesler, Mark A. Taylor, Philip J. Rasch, David C. Bader, Larry K. Berg, Peter Caldwell, Scott E. Giangrande, Richard B. Neale, Yun Qian, Laura D. Riihimaki, Charles S. Zender, Yuying Zhang, and Xue Zheng. Regionally refined test bed in E3SM atmosphere model version 1 (EAMv1) and applications for high-resolution modeling. *Geoscientific Model Development*, 12(7):2679–2706, 7 2019.
- [23] Karl E. Taylor, Ronald J. Stouffer, and Gerald A. Meehl. An overview of CMIP5 and the experiment design, 4 2012.

- [24] M A Taylor, J Edwards, S Thomas, and R Nair. A mass and energy conserving spectral element atmospheric dynamical core on the cubed-sphere grid. *Journal of Physics: Conference Series*, 78(1):012074, 7 2007.
- [25] Mark A. Taylor and Aim Fournier. A compatible and conservative spectral element method on unstructured grids. *Journal of Computational Physics*, 229(17):5879–5895, 8 2010.
- [26] Paul A. Ullrich and Mark A. Taylor. Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part I. *Monthly Weather Review*, 143(6):2419–2440, 6 2015.
- [27] Shaocheng Xie, Wuyin Lin, Philip J. Rasch, Po-Lun Ma, Richard Neale, Vincent E. Larson, Yun Qian, Peter A. Bogenschutz, Peter Caldwell, Philip Cameron-Smith, Jean-Christophe Go-laz, Salil Mahajan, Balwinder Singh, Qi Tang, Hailong Wang, Jin-Ho Yoon, Kai Zhang, and Yuying Zhang. Understanding Cloud and Convective Characteristics in Version 1 of the E3SM Atmosphere Model. *Journal of Advances in Modeling Earth Systems*, 10(10):2618–2644, 10 2018.

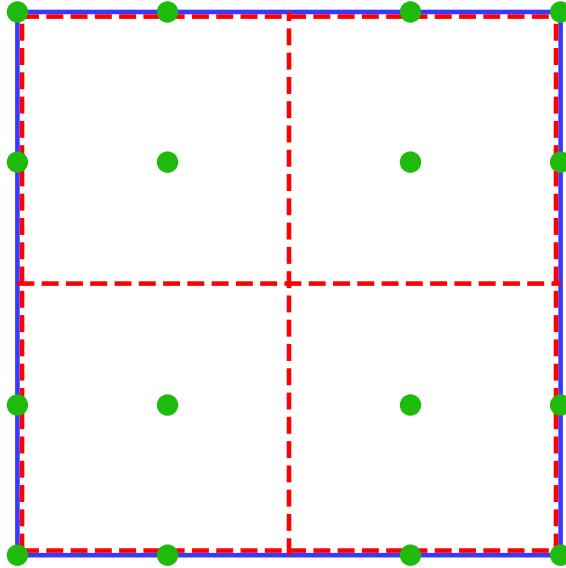


Figure 1: Schematic of the  $n_p = 4$ ,  $n_f = 2$  element. The blue solid line outlines the spectral element, the green filled circles mark GLL points, and the red dashed lines outline FV subcells.

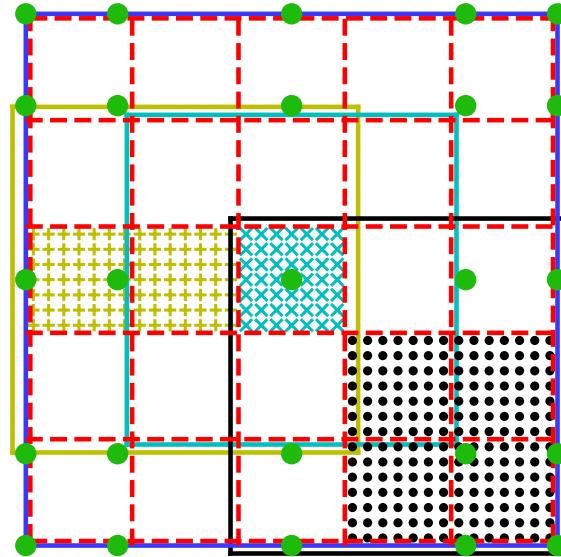


Figure 2: Schematic of the panels used in the panel reconstruction method, for an element with  $n_p = n_f = 5$  and panels with  $n_{f'} = 3$ . As in Figure 1, the blue solid line outlines the spectral element, the green filled circles mark GLL points, and the red dashed lines outline FV subcells. Filled subcells of a particular color are associated with the  $3 \times 3$  panel outlined by the same color.

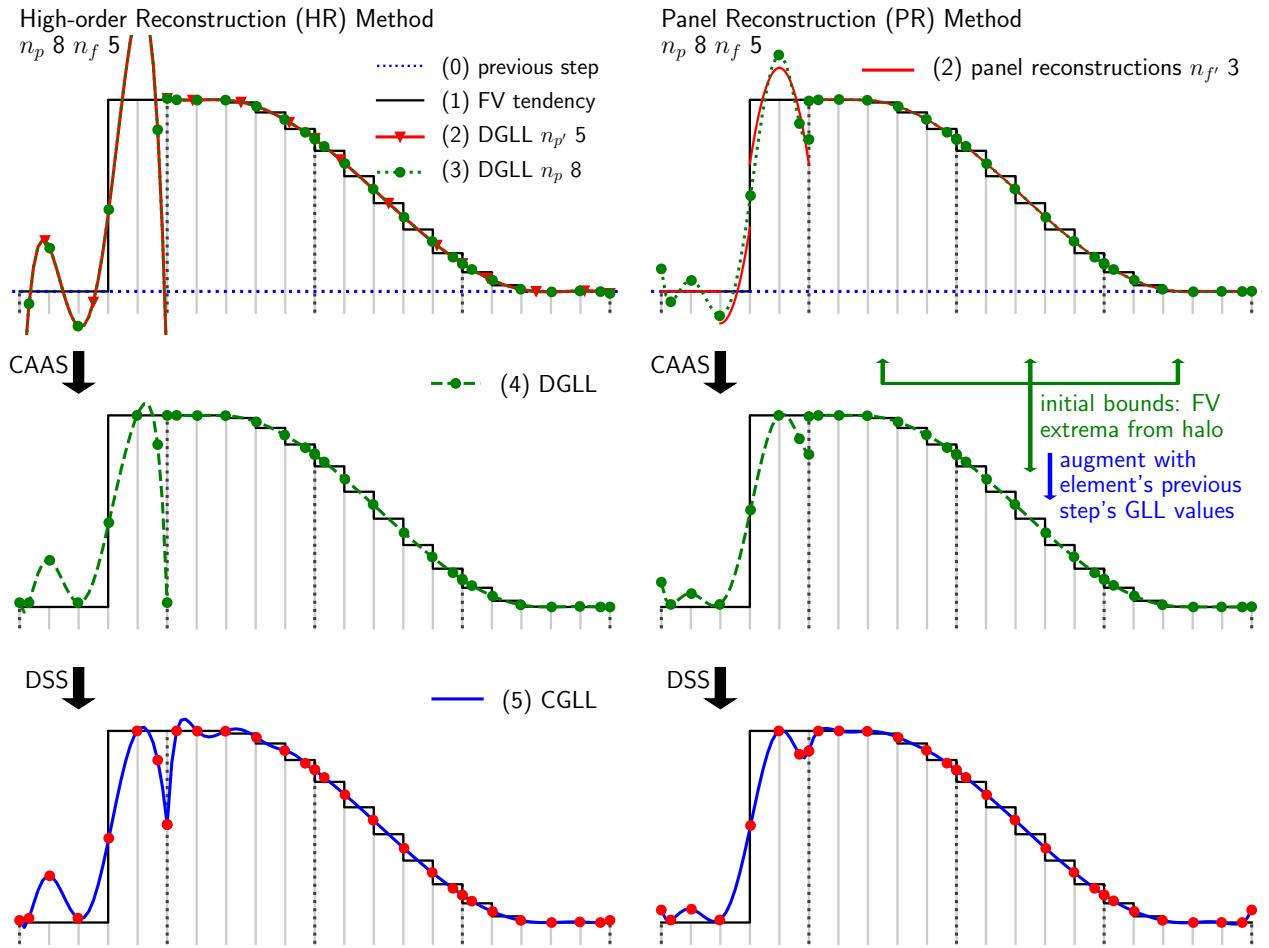


Figure 3: Schematic of the FV-to-GLL remap algorithms on the global grids, using the high-order reconstruction (HR) method (left column) and panel reconstruction (PR) method (right column). See text for details.

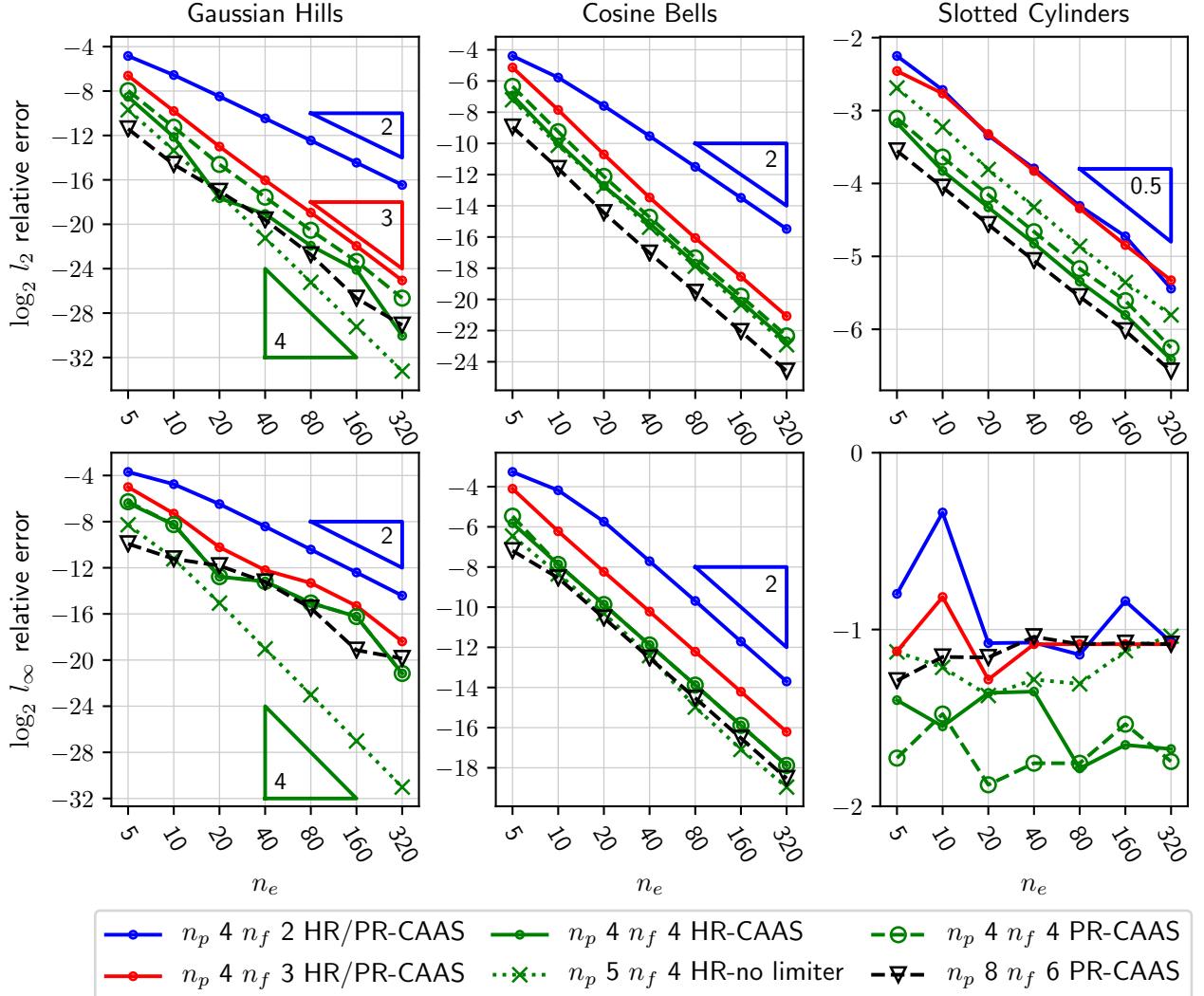


Figure 4: Convergence study of the remap algorithms. Each column corresponds to a tracer field, whose name is in the column title. The top row shows  $l_2$  error; the bottom,  $l_\infty$ . The  $x$  axis is  $n_e$ , the cubed-sphere grid element parameter. The  $y$ -axis is  $\log_2$  of the relative error. Triangles provide reference order-of-accuracy slopes. The legend provides line decoration details for  $n_p$ ,  $n_f$ , HR or PR method, and use of a limiter.

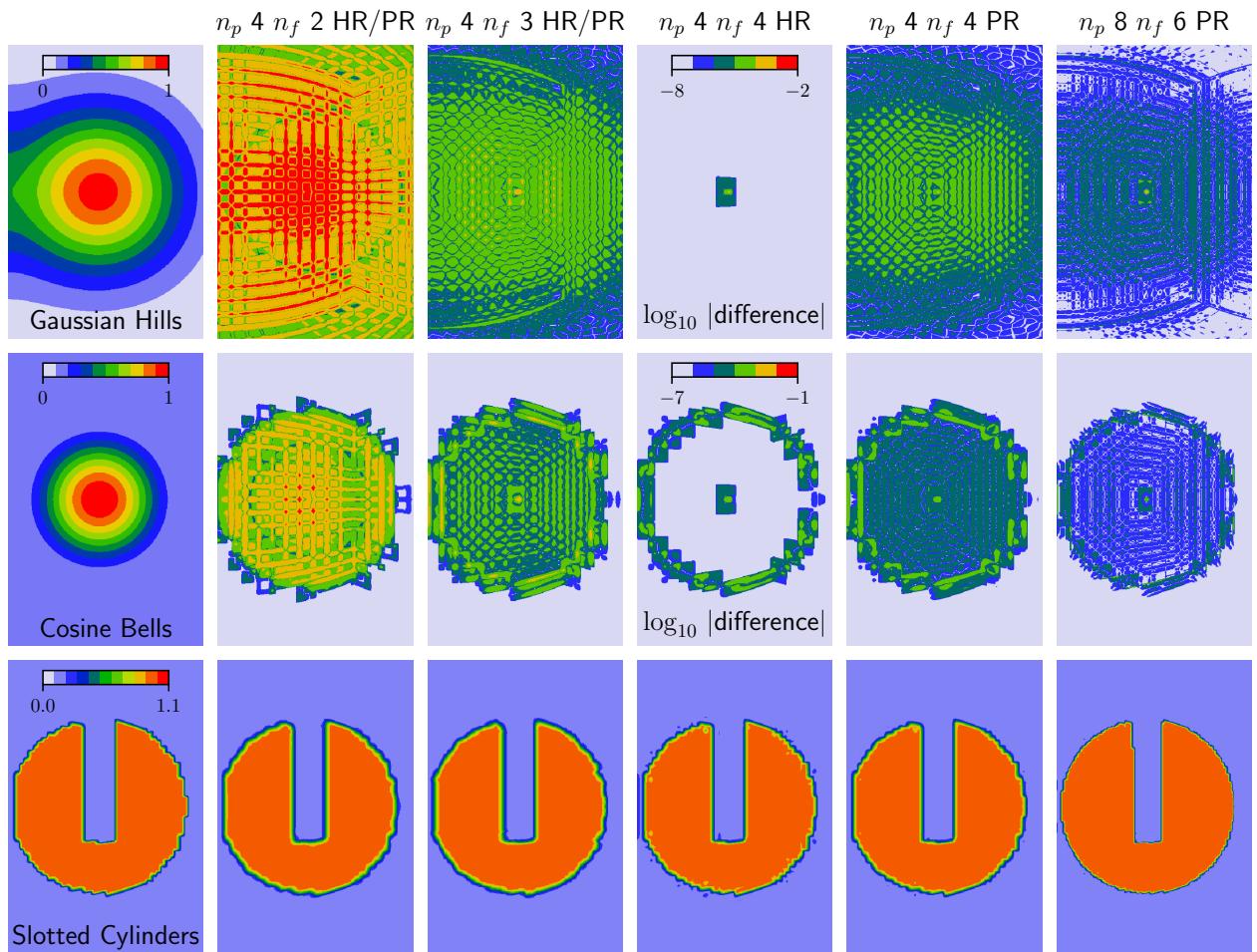


Figure 5: Contour plots for a subset of cases in Figure 4. The first column shows the original field on the  $n_e = 20$  GLL grid. Subsequent columns show the field after being remapped from GLL to FV and then back to GLL grids. Further details are in the text.

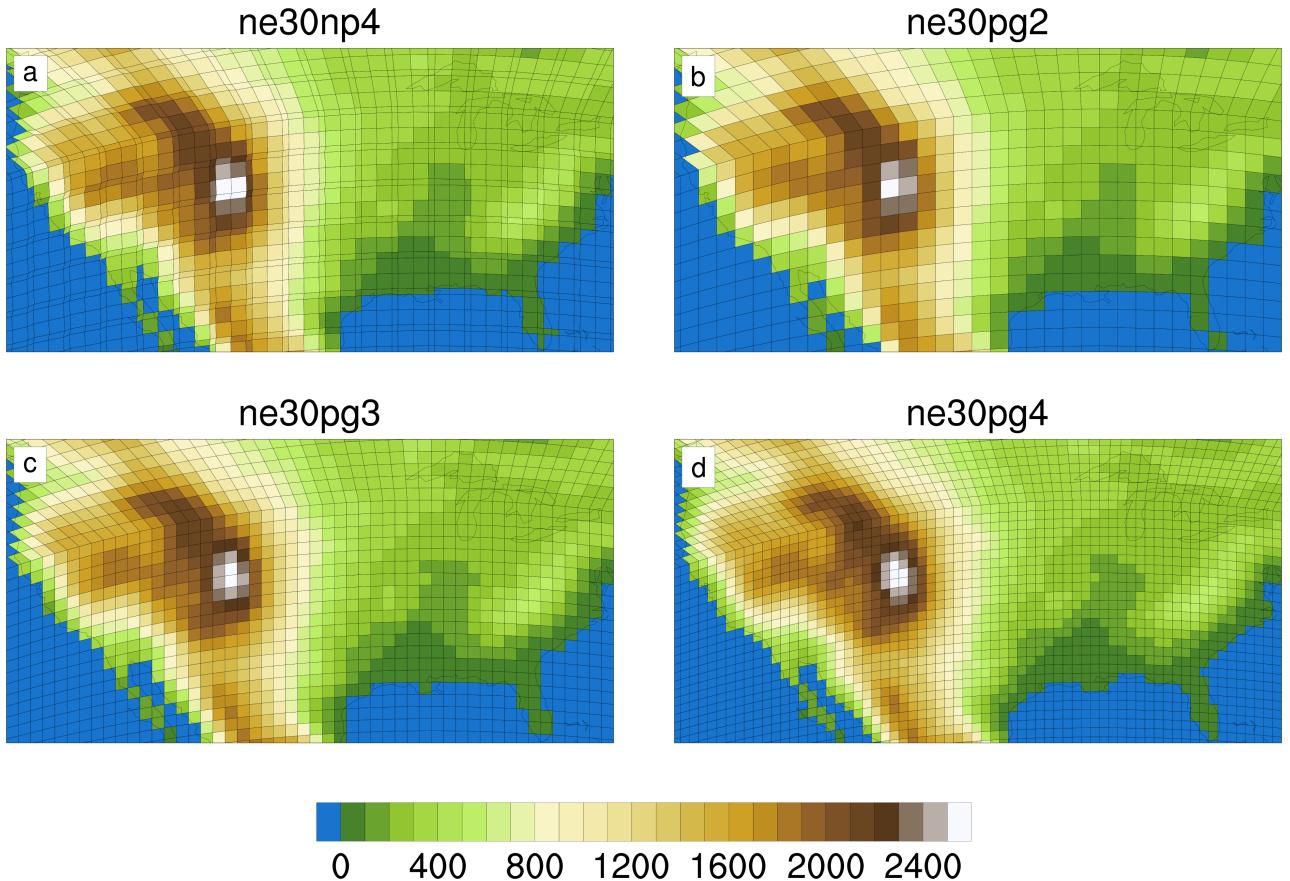
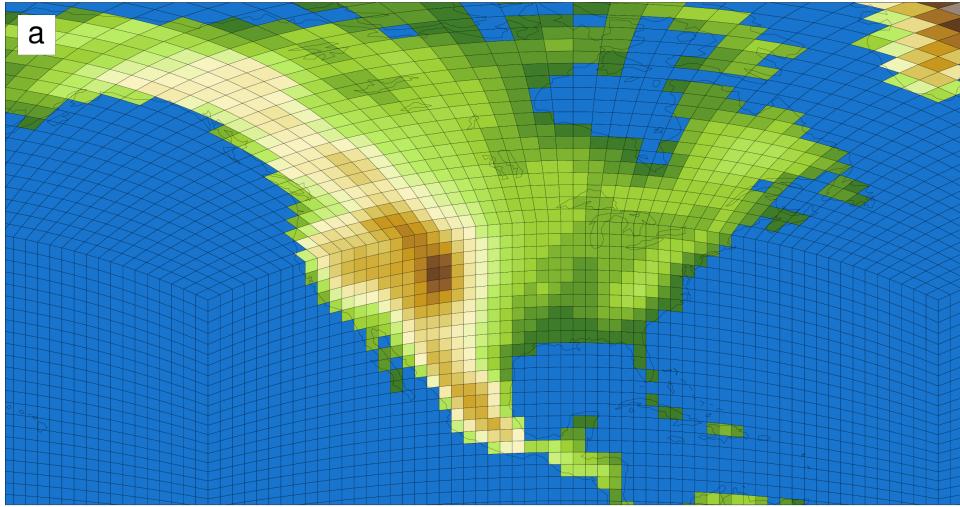


Figure 6: Comparison of physics grids used for simulations colored according to the topography of each physics column. Cells with less than 50% land fraction are colored blue. The ne30np4 or “GLL” grid (a) is represented by defining control volumes around the GLL quadrature nodes according to their quadrature weight. The finite volume physgrid can be run with  $2 \times 2$  (b),  $3 \times 3$  (c), or  $4 \times 4$  (d) cells per element.

ne30pg2



RRM pg2

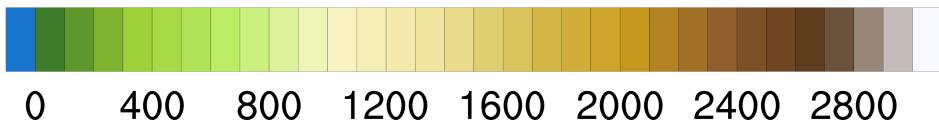
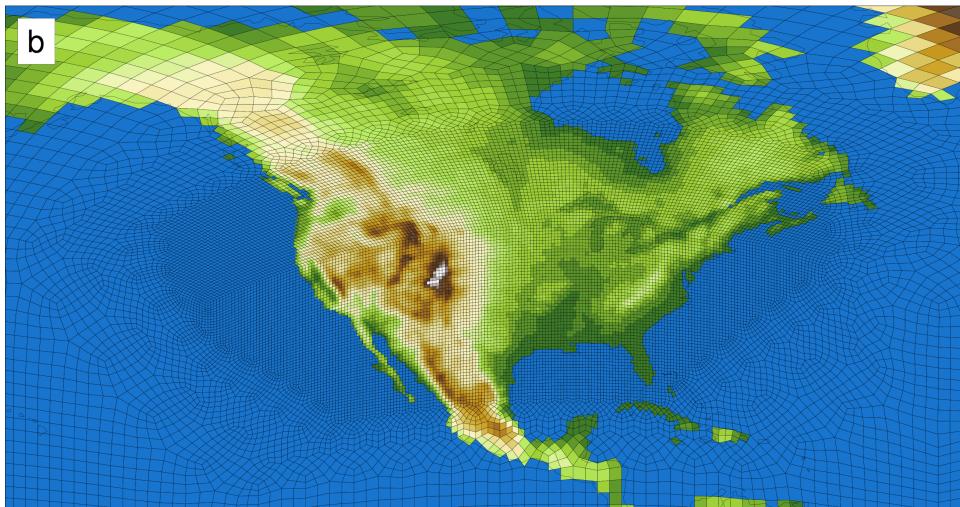


Figure 7: Similar to Figure 6, surface topography of the ne30pg2 grid (a) and the regionally refined pg2 grid (b) over the continental United States. The regionally refined grid is identical to ne30pg2 outside the refined region with a grid spacing of roughly  $1.5^\circ$ , and the refined region consists of grid cells spaced by roughly  $3/8^\circ$

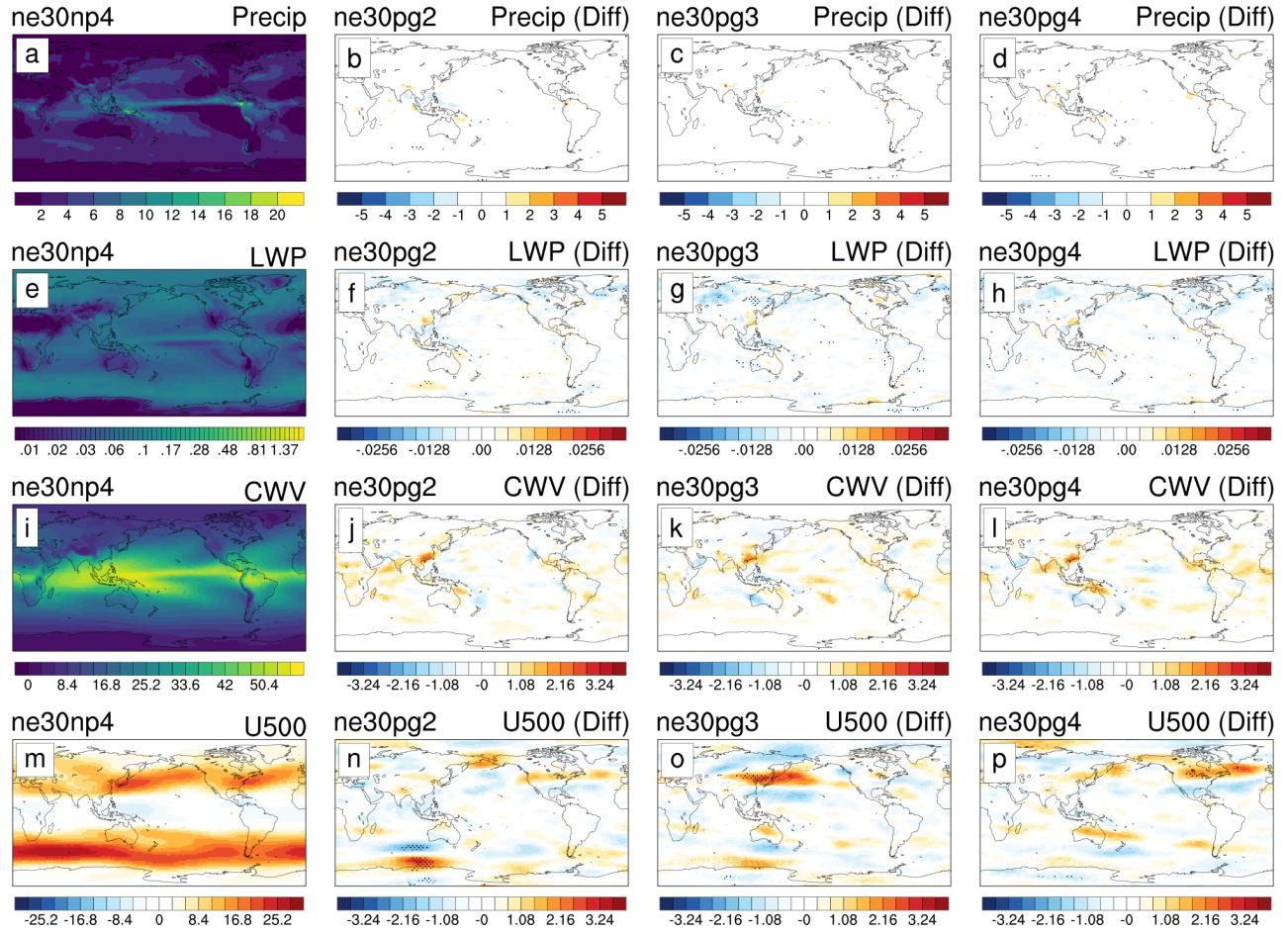


Figure 8: All season 5 year mean precipitation (top), cloud liquid water path (LWP, upper middle), and column water vapor (CWV, lower middle), and 500mb zonal wind (U500, bottom) for the ne30np4 case and physgrid cases shown as differences from the ne30np4 case. All data was regridded to a common  $1^{\circ}$  grid before averaging. Stippling indicates where the differences are statistically significant at the 95% confidence level using a Student's *t*-test.

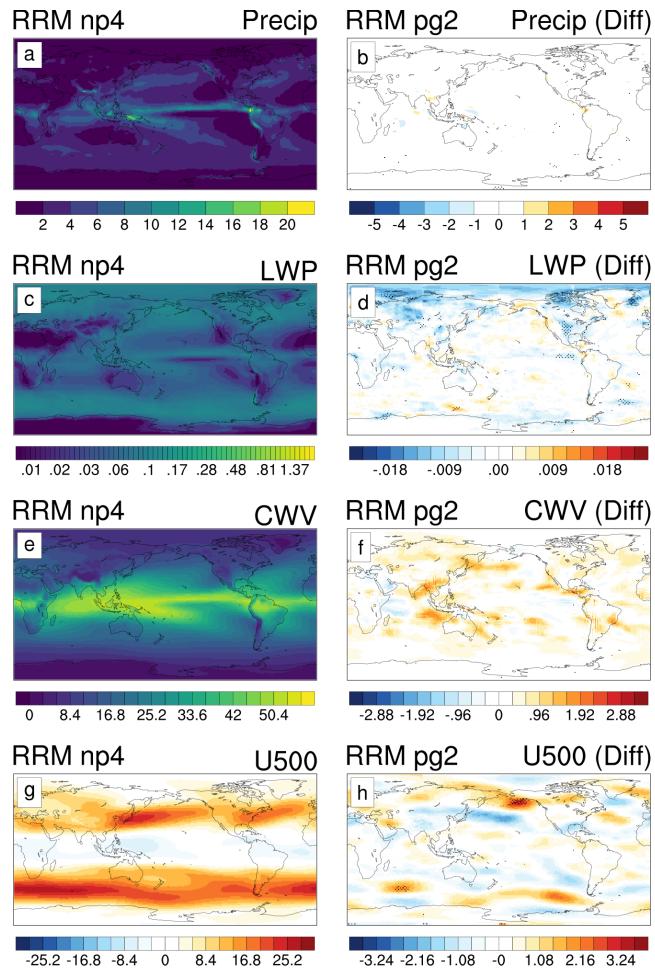


Figure 9: Similar to Figure 8 but for the simulations with regional refinement over the continental US.

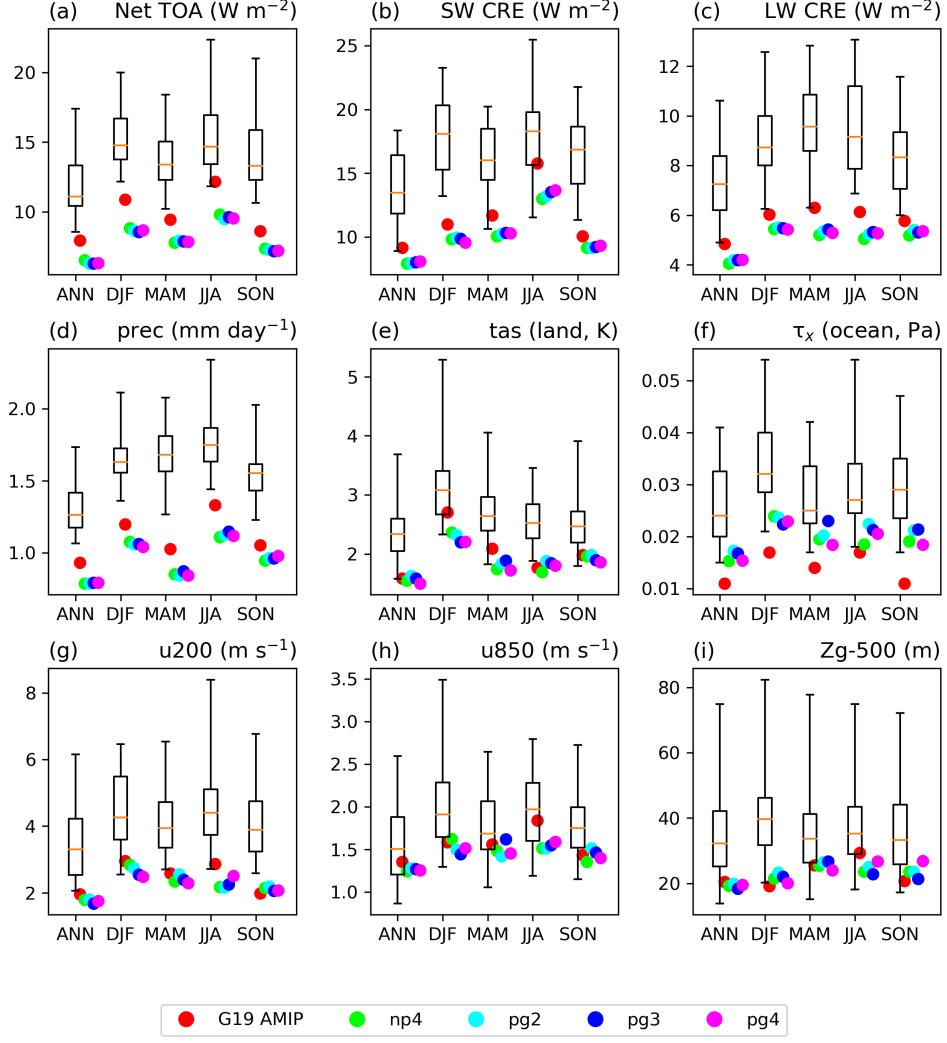


Figure 10: Comparison of root-mean-squared error (RMSE) relative to observations (1981–2005) of an ensemble of 45 Coupled Model Intercomparison Project Phase 5 models (box and whiskers showing 25<sup>th</sup> and 75<sup>th</sup> percentiles, minimum and maximum) with an E3SM AMIP simulation from G19 and prescribed SST simulations with E3SM with and without the physgrid (colored dots). Fields shown include TOA net radiation (a), TOA SW and LW cloud radiative effects (b, c), precipitation (d), surface air temperature over land (e), zonal wind stress over ocean (f), 200 and 850 hPa zonal wind (g, h), and 500 hPa geopotential height (i). TOA = top of atmosphere; SW = shortwave; CRE = cloud radiative effects; LW = longwave.

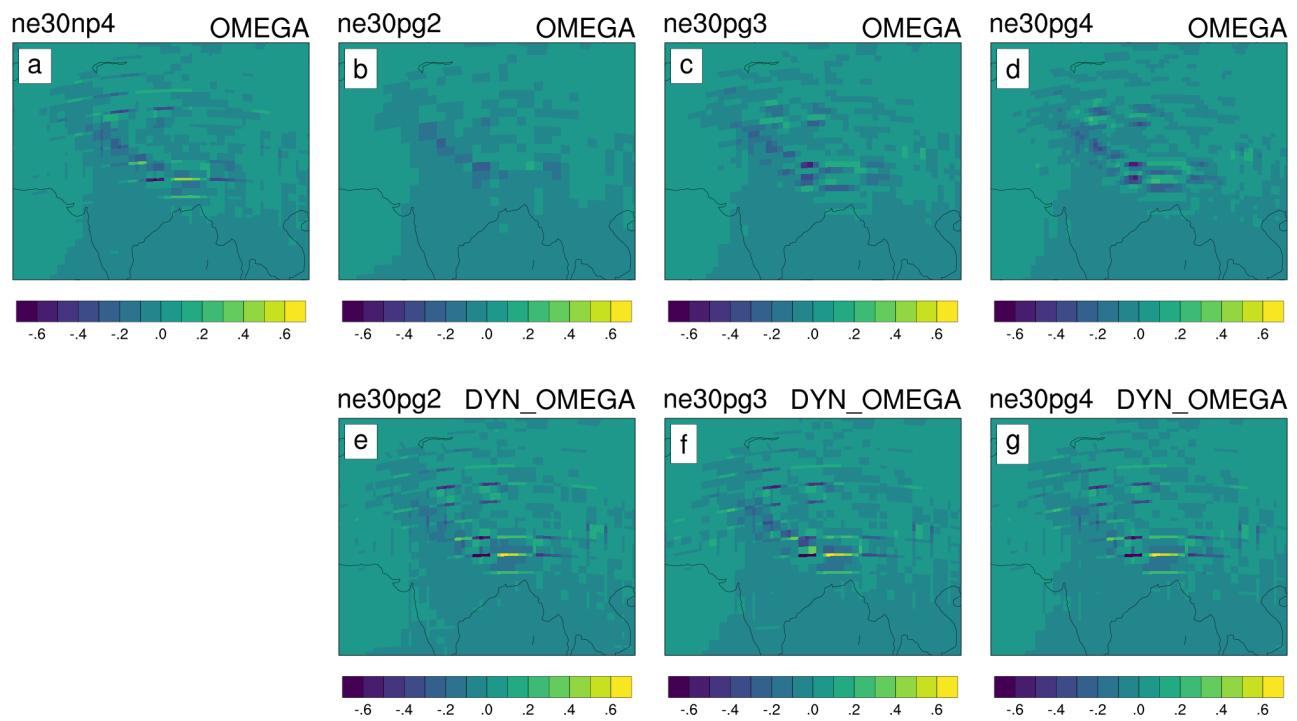


Figure 11: Maps of 5 year mean 500 mb vertical pressure velocity on the physics (a-d) and dynamics (e-g) grids.

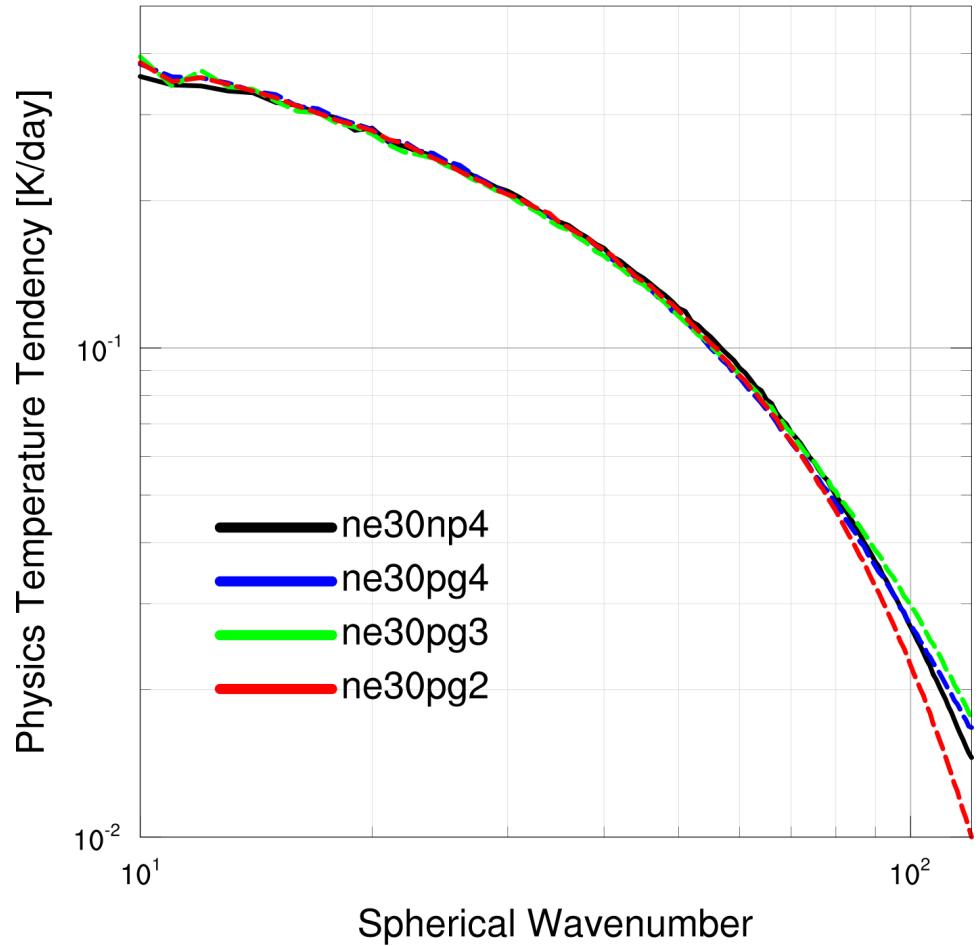


Figure 12: Power spectra of physics heating tendency calculated on a model level near 500 mb using dynamics grid data from ne30np4 (solid black) and ne30pg2, ne30pg3, and ne30pg4 (dashed) cases.

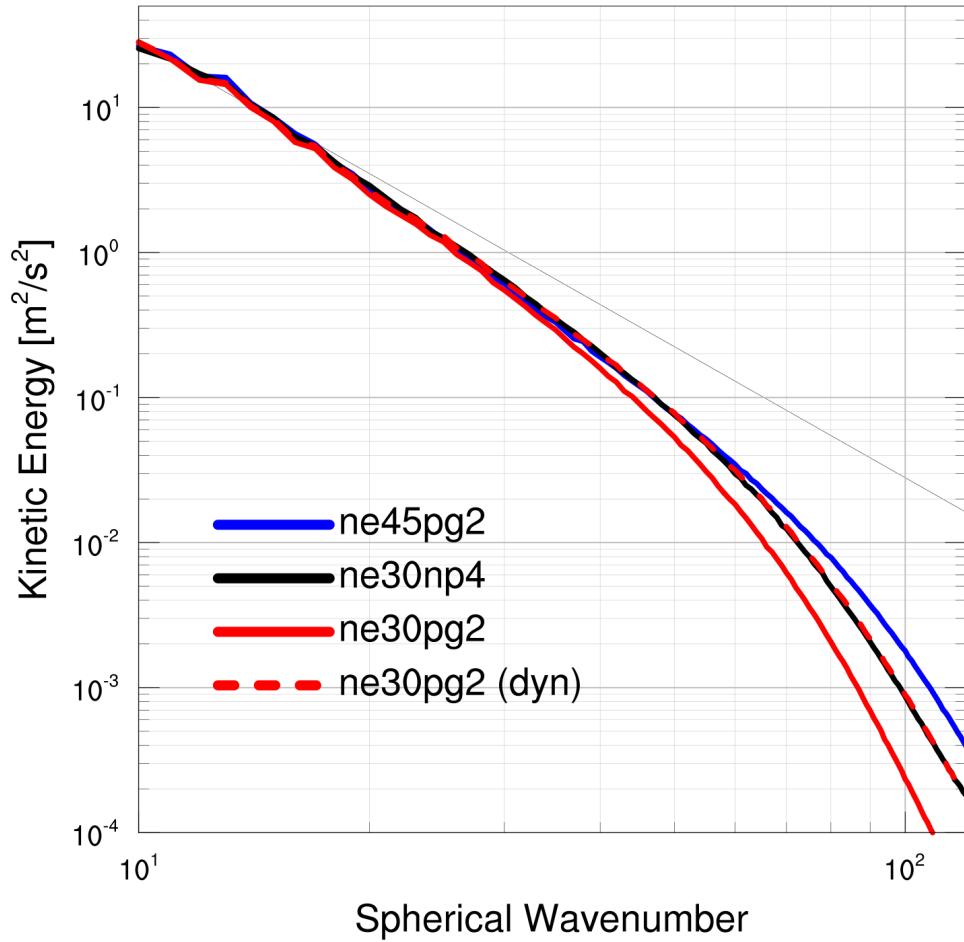


Figure 13: Kinetic energy power spectra calculated on a model level around 200 mb using physgrid data from ne30np4, ne30pg2, and ne45pg2 cases and dynamics grid data from the ne30pg2 case. A  $k^{-3}$  power law is depicted by the thin black line for reference.

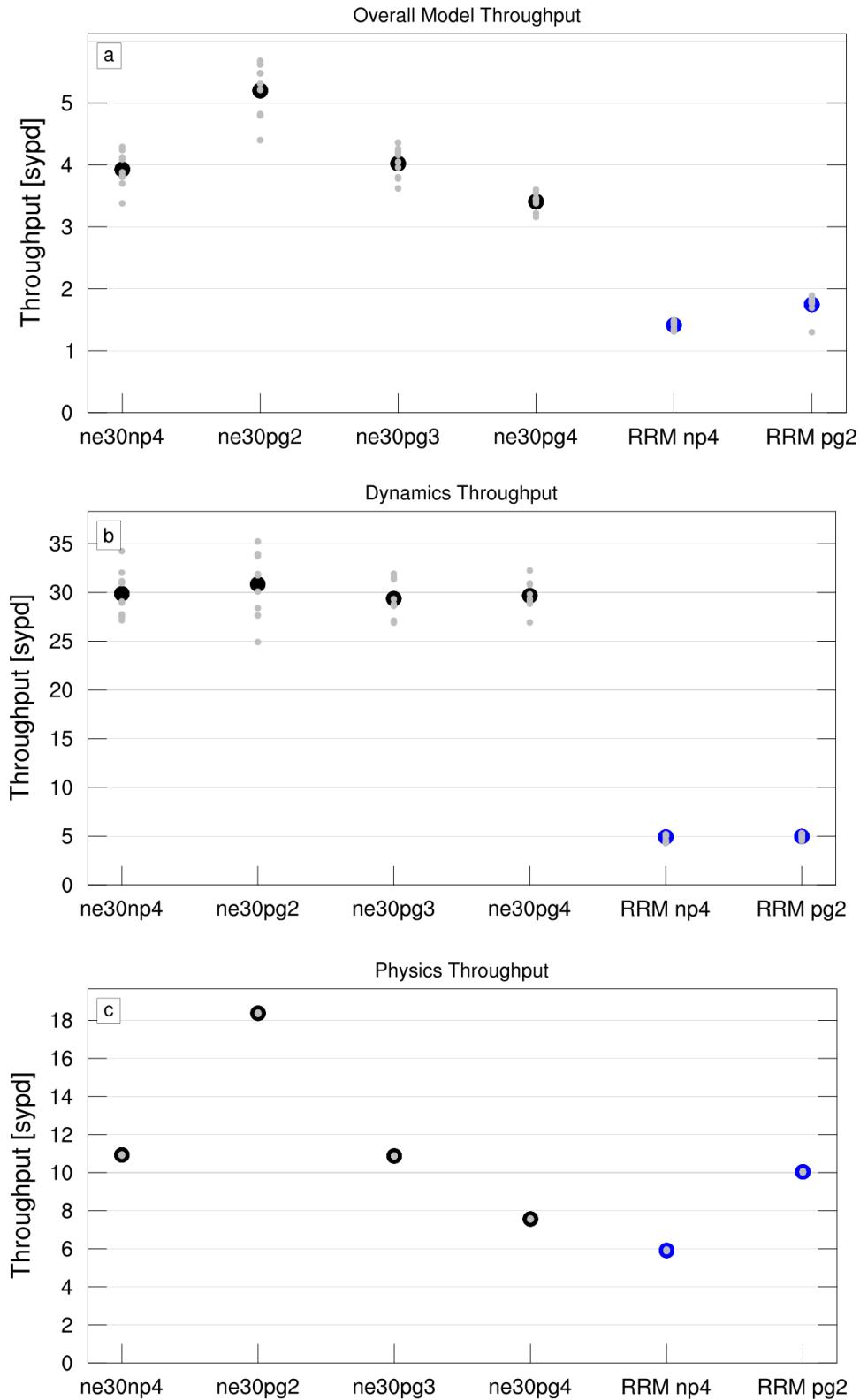


Figure 14: Model throughput in simulated years per wall-clock day (sypd) of the entire atmosphere component model (a), the dynamics calculations (b), and the physics calculations (c). The data for each individual batch submission is shown in gray markers. The average throughput over the entire simulation is shown with black markers for the ne30 case and blue markers for the RRM cases.

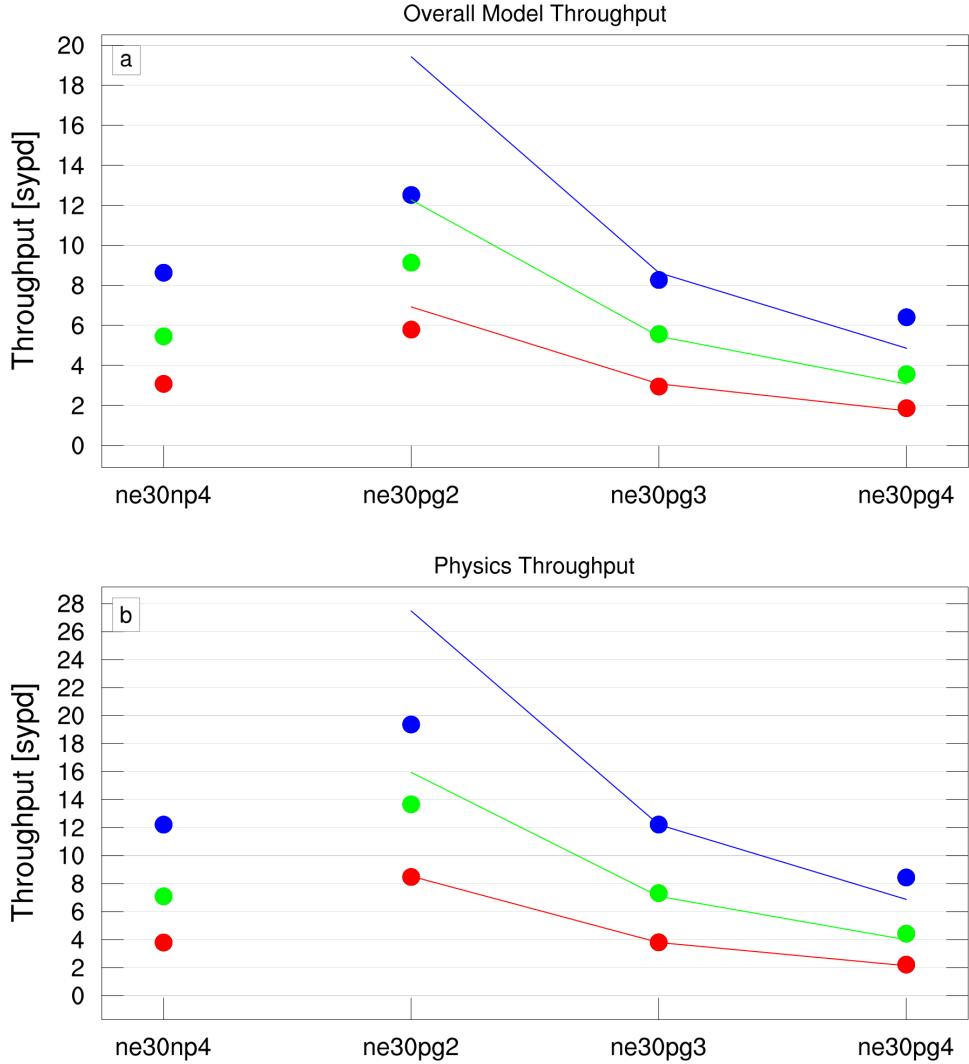


Figure 15: Similar to Figure 14(a) and (c) except for aquaplanet simulations with file output disabled. Each marker on the plot is an average of three 5-day simulations. Marker color indicates the number of Cori-KNL nodes as 22 (red), 43 (green), or 85 (blue). Lines indicate the ideal throughput based on scaling the corresponding ne30np4 result by the change in the number of physics columns.