Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-280-AC5, 2019

Geoscientific Model Development Discussions

Interactive comment on "Improving climate model coupling through a complete mesh representation: a case study with E3SM (v1) and MOAB (v5.x)" *by* Vijay S. Mahadevan et al.

Vijay S. Mahadevan et al.

mahadevan@anl.gov

Received and published: 5 March 2019

Dear Reviewer,

Once again, we thank you for the extensive review comment on the manuscript. Please find our detailed responses for all the issues raised in your review comments. We have included most of the suggested changes in the manuscript, and will be uploading the latest copy to the website. Rebuttals for specific questions have also been included in the response here and we hope that it will provide better context in light of the new changes added.

1 Needed clarifications

- Referee: p.4 l.29 explain "consistently respecting the underlying discretization" or remove the sentence.
  Authors: Modified. Please review.

- Referee: p.6 l.6 define (or cite) "component architecture". Versus what?
  Authors: The following reference has been added.
  Zhou, S. J. "Coupling climate models with the earth system modeling framework and the common component architecture." Concurrency and Computation: Practice and Experience 18.2 (2006): 203-213.

- Referee: p.6 l.21 does "Fig. 1 (right)" apply to OASIS3-MCT also? Or does the mere-library approach (no separate N_x for the coupler) defines another workflow?
  Authors: Yes. We could have $N_x = N_{atm}$ and still have Fig. (1) (left), the hub-and-spoke model, just to be clear. But what the distributed coupled model refers to is that the components can directly communicate with each other without an additional hop through the coupler (hence no explicit N_x). This is much more efficient in terms of total reduced data-transfers and optimizations that can be performed on pair-wise meshes, which are not available on a many-to-many scenario through a global coupler. In the workflow we have defined in the manuscript, the MBTR workflow allows both.

- Referee: p.7 l.25 what does "field [...] aware" mean?
  Authors: Being field aware indicates that regridder needs to understand the discretization types. Our aim is to provide an online remapper that supports consistent and conservative projection of **FV, cGLL, dGLL** source field data to target meshes, and fill the gap with ESMF based offline remapping workflows that require dual meshes and only support FV-type discretizations.

- Referee: p.7 l.29 "during the setup phase" is in contradiction with the aim of allowing for adaptive and moving meshes. Indicate whether it is just a practical choice in the current implementation.
  Authors: In the current implementation, we are computing the remapping operators only in the setup phase since most existing E3SM workflows do not support adaptive grids. But the MBTR workflow can fully support remapping with moving meshes by recomputing the weight matrices at run-time. The text has been slightly rephrased to clarify this.

- Referee: p.7 ll.30-32 in what exactly is the MBTR stack an improvement w.r.t. the MCT view, since MCT is able to handle decomposed meshes?
  Authors: MBTR is an improvement because it also stores the connectivity of the mesh. In particular, MOAB knows that the neighbor of a point might be on another processor. MCT does not have any of that information, which is essential when performing online remapping computation or performance optimizations/load rebalancing based on mesh topology.

- Referee: p.8 l.3 indicate under which scheduling assumptions the N_x processes can share with the $N_{c,l}$ processes part of the processor resources as implied later by Fig. 4.
  Authors: Sharing resources would be appropriate when the physics of the system requires that a calculation performed in the coupler must happen before the solver in the next component is invoked. Figure 4 doesn't indicate that the coupler is sometimes invoked multiple times as individual components are executed, and it specifically doesn't show the global "driver" layer which controls the overall flow of execution and data transfer.

- Referee: p.8 l.14 please define a "DoF": since it is not a word used for cell-centered couplers, it is not a common term for all the readers.
  Authors: DoF has been expanded.

- Referee: p.8 whole section 3.1 (and following) please include references or links for HOMME, MPAS, VisIt and in general do so for all mentioned models, libraries and other software tools (Zoltan, ParMetis, Eigen3, etc).
  Authors: These references have now been added.

- Referee: p.8 l.26 why "replicated" meshes. Isn't it rather "partitioned"?
  Authors: Corrected.

- Referee: p.8 l.29 "in terms of a 'Tag'." is a useless statement unless you make it clear to the reader.
  Authors: Small description of a tag has been added.

- Referee: p.8 l.29 n_p has not been defined and is not trivial.
  Authors: This sentence has been rephrased

- Referee: p.9 Algorithm 1. The formulation is too compact and missing some previous definition. Insert references to following sections for details.
  Authors: Appropriate references to other sections have now been added.

- Referee: p.11 ll.2-3 state here (or anticipate) the rational for replacing MCT as a broker.
  Authors: We expect MOAB to perform data transfers faster an more efficiently (fewer overall messages) then MCT at scale because of MOAB's crystal router. MOAB will also have better memory scaling because, unlike MCT, it does not have datatypes that can grow with grid or processor size. Finally MOAB will allow a simplified workflow by removing the need for directories of mapping weight files.

- Referee: p.12 l.8 Kd-tree is a relatively common technique (already mentioned at p.4) BVH-tree deserves a reference here (only provided at p.20).
  Authors: Added references for both tree structures

- Referee: p.12 l.11 why "unique" ?
  Authors: Removed. It is clearer now.

- Referee: p.12 l.17 the same consideration as for p.7 l.29 applies.
  Authors: Removed reference to the setup phase.

- Referee: p.12 ll.26-29 Fig 6. is not immediate to read without some further "step to step" details in the text.
  Authors: This comment is unclear. Should the advancing front algorithm be explained better ? We have added references to the front intersection video illustrations that are added as supplementary materials.

- Referee: pp.12-13 subsection 3.3.1 does the seed determination can be fully automated or its efficiency depend on user tuning?
  Authors: The determination is fully automated. However, there may be cases with failures when dealing with meshes with holes where a seed in say an atmosphere mesh may not be able to find a corresponding element containing point in the MPAS mesh (if it falls in a land geographical area). Such cases could require more than one attempt in each partition to get the front computation started.

- Referee: p.13 l.11 what does the sentence "without approximations" refer to (especially w.r.t what alternative)?
  Authors: The sentence "without approximation" refers to the fact that the intersection can be computed to machine precision as the edges become straight lines in a gnomonic plane (projected from great circle arcs on a sphere). If curves on a sphere are not great circle arcs (splines, for example), the intersection between those curves has to be computed using some nonlinear iterations such as Newton Raphson for example (depending on the representation of the curve).

  We wanted to indicate in the mauscript that intersection in gnomonic plane is simple to do and "exact"; When you have more general curves on a sphere, you

might even have multiple points of intersections, which could test the robustness and stability of the intersection algorithm. However, note that a latitude arc can intersect a great circle arc in 2 places (this can still be computed exactly to machine precision, without any approximations coming from an iteration).

- Referee: p.14 l.6 computing a meaningful bounding box is not trivial in polar or periodicity regions for lon/lat grids.
  Authors: MOAB stores explicit 3-d bounding boxes, since it is a general mesh query/manipulation library. We have not particularly encountered difficulties in handling lat/lon grids.

- Referee: p.14 l.7 does "to all tasks" refer to tasks (or rather processes) on the source side?
  Authors: This refers to processes on the coupler processing elements. The source/target meshes are already in coupler PEs, and a coverage mesh is computed by appropriately moving only elements required to completely cover target elements in current process.

- Referee: p.14 l.8 "Cells [...] are sent": how are they represented? What's the size of the communications? Is any packing strategy used to avoid latency in separate small communications?
  Authors: MOAB utilizes the aggregated crystal router to efficiently send small data between processes. In an all-to-all communication strategy, with log(N) steps of communication, all the processes get access to the data they need. This is used once during the setup phase to establish point-to-point communication links, which is then used later to pack and send data directly.

  During the field transfer from components to coupler, we pack multiple fields together in a single array to send the data to coupler, apply weight matrices on the vectors and transmit back the fields (in a packed and aggregated fashion) to the

target component. The size of such communication is on the order of DoFs in source + target.

- Referee: p.14 l.10 please clarify the term "superset": a superset usually refers to inclusion of similat objects. Does it imply that the after representation in MOAB - through the definition of the supermesh - the source and the target side share the same spatial discretisation?
  Authors: The MOAB view of the supermesh includes the union of all vertices and (elements formed by) edges in both source and target grids. Hence the supermesh is typically the superset of either the source or the target grid. This is only with respect to the actual topology of the grid, and has no correlation to the underlying discretization of field data.

- Referee: p.14 l.14 is the "crystal" router explained in Tautges et al. (2012) [N.B. reference not freely available] or does it need an extra reference?
  Authors: Added.

- Referee: p.15 l.5 how expensive can be the communication of ghost intersection elements on highly distributed components?
  Authors: The communication is typically among nearest neighbors and requires 1-2 rings of elements on average depending on the relative resolution between source and target grids. Since these are direct nearest neighbor computations that are performed only once during the setup phase, the actual impact on overall runtime is small.

- Referee: p.15 l.13 does "has the potential to" mean that is just an idea or is there a prototype?
  Authors: This is an idea and a work in progress at the moment. We expect to spend more time hardening the implementation in the coming year.

- Referee: p.16 l.28 "it is non-trivial to": did you find a way?

  Authors: There are ways that could ensure bit-for-bit reproducibility at the cost of heavy sub-optimizations. There are internal discussions to better understand whether the non-BFB algorithmic parts can be isolated together.

- Referee: p.21 l.8 reference to NE11 configuration not known to the reader.
  Authors: NE refers to the number of elements on an edge of a cubed-sphere grid. This has been clarified in the manuscript.

- Referee: p.26 Fig.14(b) provide an explanation for the difference of behaviour when going beyond 64 processors
  Authors: This was an interesting transition in the communication timings as we expanded from intra-node to inter-node regime on Cori that has 64 Haswell cores per node. The overall message passing latency as we cross the 64-core barrier is certainly evident in the figure, especially since we are only communicating one solution data field from the component to coupler and vice-versa. We have added some text in the revised paper to discuss this further.

2 Technical Corrections

- Referee: p.2 l.5 vs l.31 (and elsewhere) make the use of "donor" or "source" consistent p.2 l.9 should probably be "conservation for critical quantities" p.3 l.20 the subject of "that nonlinearly couple" should not be "solution fields" (they are just exchanged in the nonlinear coupling process)
  Authors: Done

- Referee: p.6 l.6 unclear (if not useless) reference "Section (1)"
  Authors: Introduced a subsection 1.1 to clarify.

- Referee: p.6 l.32 the "GLL acronym" is used before definition which is given a

few lines later
Authors: Done

- Referee: p.7 l.29 "an in-memory" instead of "a in-memory" p.8 l.26 remove "a" before "replicated SE and MPAS" meshes
  Authors: Removed "a"

- Referee: p.9 Algorithm 1. - Step 1: if l can only be s or t - as in step 4: - indicate $l \in [s, t]$ also in step 1: otherwise if the formulation is generic for more than one mesh for component, the naming should be consistent. - Step 2: if you indicate W_{ij} instead of W_{st} you should not define i,j as a mesh pair. Later at step 20: i takes a specific meaning.
  Authors: Revised and included suggested changes

- Referee: p.12 l.1 "partitioner" instead of "repartioner" p.12 l.23 remove "is" before "results" p.12 l.26 "each" instead of "Each" p.14 Fig.7 caption: "fully covers" instead of "fully cover" p.15 l.3 "the intersection vertices [...] need" instead of "needs"
  Authors: Done

We request you to review the updated paper when it becomes available, and we welcome any further comments that would improve the scope of the manuscript. Best regards,

Vijay Mahadevan

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-280, 2018.