- @ line 50 Rephrase this to reflect what is accomplished, and to distinguish it from future plans.
  - We've rewrote the paragraph to be more specific about the differences between FEOTS and other capabilities, and to be clear about the current state of FEOTS and what we plan to do in the future.

- @ 62 Convergence rates? Perhaps you mean volume transport rates? (So as not to be confused with the convergence rates to an equilibrium solution.)
  - Here we were talking about the error convergence rate of the time integration scheme. For example, the Forward Euler method exhibits linear convergence with the time step size. When we were drafting this paper, we initially had plans to look into convergence rates. We ran into a few issues when looking at this and opted to leave this out of the paper. To avoid any confusion, we've removed the mention of convergence rates at this line.

- @ 93 Section Graph Coloring approach to Operator Diagnosis This would be easier to comprehend after having read Section 2.4 Parent Model, where the advection scheme is discussed, and the motivation for the graph coloring approach becomes clear.
  - We have moved the previously named "section 2.4" to occur immediately before this section; Reviewer #2 also made this request.

- @ 125 State which advection scheme (Table 1) was actually used in the offline transport matrix model for the example in this paper.
  - We have added the following to the caption on Table 1

    3rd Order Adams-Bashforth is used for the Argentine Basin test problem presented in this paper.

  - We have added the following to the end of the "Time Integration" section

    For the results presented in this paper, we use the 3rd Order Adams-Bashforth time integrator and the conjugate gradient solver is stopped when the residual magnitude, relative to the initial solution guess magnitude, is less than 10^{-6}.

- @ 150 and forward Somewhere (Section 2.4?) please indicate the timestep used, and the relationship to the CFL for both the parent and the transport matrix model.
  - We diagnosed the CFL values from the standard POP diagnostic output, and calculated the maximum CFL by performing an eigenvalue analysis on a transport matrix. The latter procedure was performed using matlab on a regional operator for the Argentine Basin, as memory limitations prevented us from doing so on a global operator. The dominant eigenvalue indicated a CFL value of 0.1 for a 15-minute time step.

○ We added the following text to the "Parent Model" section:

With a time step of 7 minutes, the model typically yields maximum CFL values of $\textbf{O}(10^{-1})$ or smaller.

○ We added the following text to the "Time Integration" section:

We use a 15 minute time step, and a typical maximum CFL value, obtained by eigenvalue analysis of the transport operators, is $\mathcal{O}(0.1)$.

- @ 164 and forward From this it is assumed that there is NO horizontal diffusion term, as none is generated by the online parent model. Please restate the offline equation (eq. 10) with the missing horizontal diffusion term, and clarify which of the options in Table 1 was used.
  ○ You are correct; there is no explicitly implemented horizontal diffusion term. Lateral tracer diffusion comes from the diffusive nature of the Flux-Limited Lax Wendroff scheme when it is applied to impulse fields. We have added the following statement immediately after Equation (4):

  Since we do not enable explicit lateral tracer diffusion in the parent model in this study, all elements of D_h are zero; we do not explicitly diffuse tracer laterally.

- @175 A 63-year spin-up seems short for examining deep water masses; the basin extends to 6000m. A comparison with observational temperature, salinity, and flow statistics as was done in Weijer et al, 2020 would be a good addition.

  ○ In this paper, we are not concerned with the deep water masses per se, but rather with the effect of the circulation on tracer distributions. After 63 years, the primary circulation features (boundary currents, the eddy field, the Zapiola Anticyclone) are well established. We added the following text to clarify:

  Although the model was run for 186 years, we diagnosed the transport operators for the 5-year period starting at simulation year 64. Even though 63 years of spin-up is not sufficient to fully equilibrate the stratification in the deep ocean, the main circulation features (e.g., boundary currents, the eddy field, the Zapiola Anticyclone) are well established by then, making this an appropriate data set to demonstrate the capability of FEOTS. We refer to \cite{Weijer2020a} for evaluation of the hydrography and circulation in the Argentine Basin in a companion simulation.

- @177 Is the data volume given for IRF-output and transport operators for the entire ocean, or for the region to be studied?

- ○ The parent model is a global climate model. The data volume is for the whole parent grid (global). We use this database as a source to generate either global or regional transport operators from. The paragraph has been moved to the "Graph Coloring" section and has been revised to read:

  For our test problems, we diagnosed the 5-day averaged IRFs and vertical diffusivities for the 5-year analysis period of the parent model. We repeated the simulation for 105 days, diagnosing 1-day averaged IRFs. With this methodology and the 7-minute time step, the one-day averaged operators are each an average of 1440 IRF snapshots and the five-day averaged operators are each an average of 7200 IRF snapshots. The data volume of the global parent model five years' worth of 5-day averaged operators (365 IRFs and diffusivities) is about 9 TB. Once transformed to transport operators, the data volume is 4 TB.
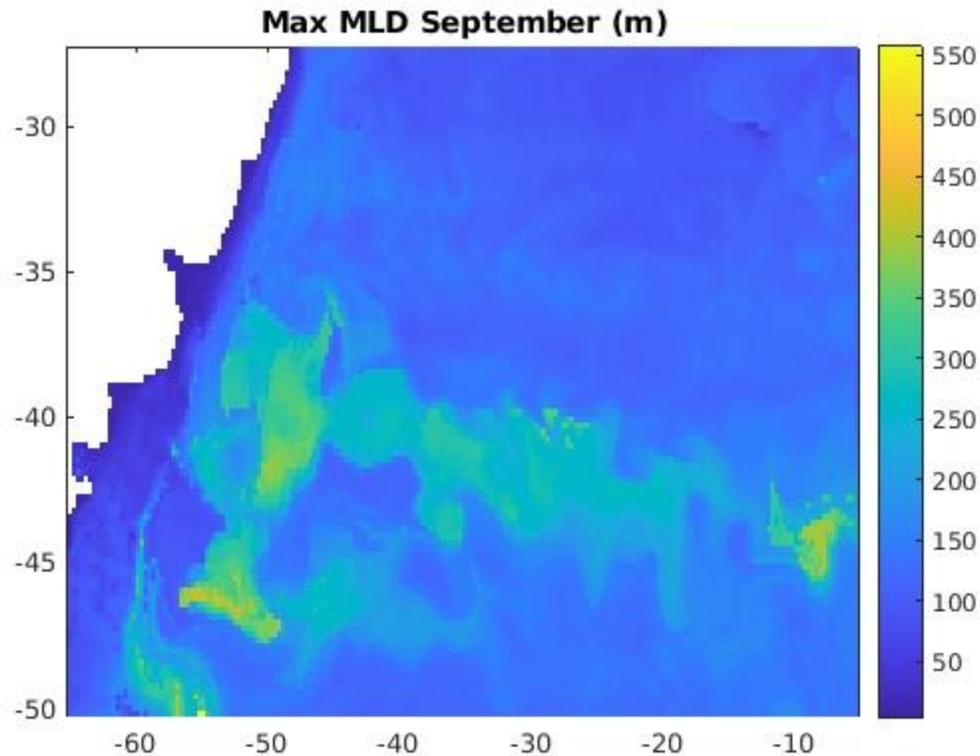
- **●** @ 178 and forward: Clearly state when you are running the global model versus the regional model.
  - ○ We've reviewed the manuscript and made adjustments following this recommendation

- **●** @ 200 and forward: This is a clear and useful discussion of the constant preservation issue, which is frequently overlooked until strange results show up. An additional useful criterion is the rate at which total tracer quantity is gained or lost. This will impact the ability to get a physically meaningful equilibrium solution.
  - ○ Looking back at this, we agree that this would be a useful metric. However, our goal when writing this section was to comment on how well constant preservation is maintained, even under single precision arithmetic and aggressive optimization. Since the manuscript was originally submitted, compute systems have changed and we have had to scrub simulation data to save on storage. To produce this metric at this point would require some time and effort that we currently do not have the funding for at the moment.

- **●** @ 203: The use of single precision for the parent model will limit the usefulness of the results. It also makes it difficult to distinguish the sources of errors in the analysis.
  - ○ The parent model is run in double precision, but the transport operators are stored as single precision. It is not clear to us why single precision should limit the usefulness of results. The Constant Preservation results provide some measure of round-off error noise in the volume field (2-3 orders of magnitude smaller than surface values after 5 years of integration). Errors in an $O(1)$ constant tracer field are at worst 3-4 orders of magnitude smaller than the expected answer (a constant field). You are correct in that we cannot decipher between numerical and round-off error based on our results; this also cannot be done if we were just running with double precision. Perhaps the comparison of single to double precision is a topic for another study.

- @212 this paragraph: Clarify that the error data is from running the global or regional version of the offline model. There appears to be an assumption that these errors are not present in the parent dynamic model. It would be instructive to run this test on the dynamic model, in order to compare the error level inherent in the parent dynamic model.
  - The discretization for tracers in POP and in the offline volume preserve constant tracers. We've added an theoretical analysis in this Methodology section that shows a constant initial condition will remain a constant under the discretization when using exact arithmetic and in combination with the time integrators presented. The source of error for the constant preservation test case arises from floating point arithmetic, which we characterize in perhaps the most catastrophic scenario - single precision arithmetic with aggressive optimizations.
  - We agree that it would be useful to perform this conservation test with the parent model for comparison. Unfortunately, the machine that we used for these simulations has been decommissioned, and there are no plans to port E3SMv0-HiLAT to the new machine at LANL.

- @ 222: …"with and without mixing": vertical mixing? Please clarify.
  - The additional analysis described above now clarifies "with and without mixing". The intention of showing this in the constant preservation case is to highlight how variable coefficient vertical mixing can lead to an amplification of round-off errors above 1000m. We feel it's important to understand this behavior.

- @233 How does the 1000m division relate to the mixed-layer depth of the model? In Figures 4 and 5, Dye 4, especially, raises the question about the depth of the mixed layer in this vicinity.

  - The figure below shows the maximum mixed layer depth in the Argentine Basin in simulation year 64 during the month of September, which is the month when the mixed layer is deepest. Clearly the mixed layer is quite deep, but it does not reach as deep as 1000 m. We included the following statement in "The Argentine Basin Test Problem" section:

    Note that maximum mixed layer depths in the Argentine Basin in winter are around 500 m in this model, so deep convection should not play a role in the transport of these tracers across the 1000 m depth horizon.

Max MLD September (m)

- Figure 4: The scales on the figures are not all the same. This creates the visual impression that the small contributions are more important than they are, compared to the major contributors. Figure 5: Interesting set of graphs. Again, the scales on the graphs are very different, creating a visual that is out of proportion for the more slowly arriving contributors.

  - From an oceanographic perspective it is certainly relevant that there is an almost two orders of magnitude difference in the inventories of the 6 dye tracers, depending on their release location. But that is not the point that we are trying to make here. The purpose of Figs. 4 and 5 is to compare tracer inventories produced by the online and offline methods, and the color and axis scales were chosen to optimize the information content of these plots. Plotting these inventories on the same scale would significantly diminish -or even nullify- the value that these plots may have about the online/offline comparison for the less abundant dye tracers.

- @340 Supply a reference for the statement : …"reduces to a first order upwind scheme".
  - To address this comment, after changing the order of some of these sections, we've moved this discussion to a new section following the Graph Coloring to Operator Diagnosis section. We've added statements that describe how a TVD flux limiter works, with references, to illustrate this point.

- @361 Need to add finding the equilibrium solution under future work.
  - The conclusions paragraph has been expanded to incorporate this note.

- @368 Your conclusion that the analysis "highlights the limitations of the IRF approach", considering the limitations of this particular implementation, might be more appropriately stated as "highlights the challenges of the IRF approach".
  - This change has been made in the conclusions and abstract.

<u>Reviewer 2 Comments</u>
- @ 6 You don't say the nature of these benefits: is your method faster ? More accurate? Also, I don't know if "clearly shows" is the best phrasing, because it is not determined whether you approach is more accurate than the online approach (it seems more diffusive) and in terms of time, as far as I understood, the comparison is done on estimated costs and not actually running the online vs. offline head to head.
  - We understand your point. In other responses below, we address why online v. offline head-to-head comparison is not done and why we have chosen to rely on an estimate. The purpose of this paper is to show progress in FEOTS development as a toolkit for offline tracer simulation. This part of the abstract now reads

    The demonstration illustrates progress in developing offline passive tracer simulation capabilities, while highlighting the challenges of the Impulse Response Functions approach in capturing tracer transports by a non-linear advection scheme. Our future work will focus on improving the computational efficiency of the code to reduce time-to-solution, using different basis functions to better represent non-linear advection operators, applying FEOTS to a parent model with unstructured grids (MPAS-Ocean), and to fully implement a Newton-Krylov steady state solver.

- @ 35 I think you should spell out what the state of the art approach is and in what way your method improves it. Your narrative is a little too vague

  - We now discuss that existing TTM models have been applied to non-eddying ocean states only, and that FEOTS is specifically designed to tackle the large computational problems associated with tracer transport in a global eddying ocean.

- @ 45 Although above you said that climate models use parameterizations to describe the presence of eddies, can you clarify?

- ○ We agree that the logic of these sections did not work well, and we rewrote this section. We argue that low-resolution GCMs cannot represent processes like those observed in the Argentine Basin, and that the current generation of TTM models, which are based on low-resolution GCMs, are therefore inadequate tools to study tracer mixing in the Argentine Basin.

- @ 50 Could you expand on the current state of the art for the IRF approach and say more on what the differences with your work are ?

  - ○ We rewrote this section, also in response to the reviewer's previous comments, as we argue that the main innovations of FEOTS is its capabilities to diagnose operators using a method that requires fewer impulse fields (using graph coloring), to solve problems associated with eddying ocean states, and to provide capabilities for modeling regional subdomains of a parent model.

- @ 60 You may want to rephrase this and just say what the magnitude of the error is (like 10^-something)
  - ○ We've made the suggested change.

- @ 65 What application ?
  - ○ FEOTS ; the phrasing of these sentences felt awkward to end the previous sentence with FEOTS and start the next sentence immediately with "FEOTS" though we understand the confusion this has caused. This section has been modified to help improve clarity.

- @ 79 What discretization method are you using ? where are the unknowns located on the mesh ? what type of cells are you using ? The section on the Parent Model should be moved up so this sentence is put into context for the reader.
  - ○ The Parent model section has been moved to the first subsection of the methodology section; this was also requested by the other reviewer. We have reiterated at this location that, for our example discussed in this paper, we are using the Flux-Limited Lax-Wendroff advection scheme on POP's Arakawa B-Grid.

- @84 Why is your offline approach better than just discretizing the operators and coding them up ? What is the accuracy of this method ? Has it been tested against toy problems for which there was an analytical solution? If so, proper references should be cited here. If not, at least a heuristical discussion on accuracy should be added.
  - ○ Our approach captures the parent model discretizations along with the fluid velocity fields. Diagnosing operators with FEOTS assistance only requires knowledge of the advection and diffusion stencils used in the parent model. For linear discretizations (with respect to the tracer fields), the sparse matrix that is diagnosed is an exact representation of the advection and diffusion subroutines in the parent model. This methodology is demonstrated in the two references

cited in the highlighted sentence. When diagnosing operators from nonlinear flux-limiting/TVD advection schemes, the methodology will diagnose a more diffusive upwind advection operator than what the parent model advection scheme is capable of. A discussion of this issue has been added to the following sections
- Parent Model
- A new section following the Graph Coloring approach to Operator Diagnosis section

The methodology we have presented is meant to be functional for a range of parent models, not just POP. Given a mesh (structured or unstructured) and the computational stencil that reflects the domain of influence of the underlying discretizations, FEOTS provides the infrastructure to diagnose an existing model's advection and diffusion schemes. The "glue" between FEOTS and a parent model is made up of a code layer that reads in a parent model's mesh and translates the connectivity information into a standard undirected graph representation. Once in this form, FEOTS can map between sparse matrix and 1-D vector representations and the parent models' memory layout. This allows us to use the same set of routines for offline simulation, independent of the parent model. In our view, this significant amount of code reuse reduces code maintenance as community models (that we would use as parent models) fall in and out of popularity.

Another benefit of capturing transport operators as sparse matrices is that we will be able to lean on linear algebra packages to assist us with future development. Given that we have proof-of-concept for our OO Fortran implementation of this methodology, we clearly have some work to do on improving time-to-solution. Packages like PSBLAS will help us expose parallelism in SpMV operations as well as implement solvers that will be necessary for tracer equilibration, which we plan to implement in future releases.

We have updated the introduction to incorporate these kinds of statements.

- @ 87 Please define in mathematical terms what you mean by "tendency of the impulse function".
  - The tendency is the time rate of change of the tracer field; it is equivalent to the impulse response function. We've rephrased this sentence and the equation below to provide a mathematical definition of the impulse response function.

- @ 89 Is Eq. (14) a better way to write this ? If A is the operator you are trying to estimate and R is known data, you should probably just write it as in Eq. (14).
  - Equation (14) is a good way to write this; this section has been rewritten, incorporating your other comments and those of the other reviewer to clarify the

> connections between the impulse response function and the matrix representation of the transport operator.

- @ 100 Does this procedure depend on the time step? In the fully discretized system, the entries of A are multiplied by delta t (time-step) I believe. Do you need to estimate again A any time the delta t for your simulation changes ?
  - In short, this procedure does not depend on the time step. The changes made in response to your previous comments should clarify this.

- @ 108-109 Is this the main difference between your approach and Bardin 2014 ? Why 125 tracers and not another number ?
  - This is one difference. We also are focusing on eddy-permitting and higher resolution simulations which has also motivated developing a software package in a compiled language. In Bardin 2014, their parent model used an advection scheme that extends two grid-cells in each spatial dimension, giving a 5x5x5 brick (125 grid points). Bardin 2014, like others who have followed this methodology (e.g. Khatiwala), simply defined "checkerboard" fields for the impulse functions, using a formula like we show immediately below this comment.

    We've added the statement "... giving a 5x5x5 brick for the domain of influence." to this sentence to make it clearer why 125 tracers.

- @ 130 Who is v ? it was not in Eq. (4)
  - v is the volume anomaly. Equation (4) has been adjusted to include the volume anomaly correction. The text beneath equation set (4) now describes the volume anomaly term and why it is included (with a reference to the POP manual).

    We've also added a description for v in the Time integration section and why it is calculated the way we do. Additionally, we've added exposition in the Constant Preservation results, per the other reviewer's request that shows how this formulation conserves total tracer.

- @ 130 How are the volume anomalies discretized ? if one wanted to reproduce what you do, how do they deal with the volume anomalies ?
  - The volume anomaly is a discrete field that resides on the tracer grid points of the Arakawa B-grid. The source of volume anomalies is from the non-zero vertical velocities at the ocean free surface. The modeled tracer field, both in the "online" parent model and in the "offline" FEOTS simulation, is a tracer concentration (a per unit-volume metric). We want to be able to conserve the total amount of tracer (volume multiplied by tracer concentration. Equations 9 and 10 show how the volume anomalies are advanced in time and incorporated into the tracer equation.

- @ 149 Can you clarify the differences between this work and Bardin 2014 ? Up to this point it sounds like you are following their approach almost entirely.

- ○ The main differences between are work and Bardin 2014 are as follows
  - ■ Our focus is on providing a toolkit for offline eddy-permitting/eddy-resolving simulations of passive tracers; Bardin (2014) focused on 1 deg resolution simulations. Global simulations with FEOTS are currently slow, but this paper aims to highlight some progress we have made in the initial version of FEOTS.
  - ■ FEOTS provides a suite of compiled Fortran programs to assist in impulse function creation, sparse matrix diagnosis from the impulse and impulse response functions, and offline tracer simulation. It is written using OO principles to allow extensibility to other parent models. This is in comparison to a monolithic MATLAB implementation. We opted to go with a compiled language for performance reasons; removing JIT compiling provides performance gains; the ability to eventually parallelize this code through domain decomposition / parallel sparse matrix libraries will help us tackle global offline eddying simulations.
  - ■ Our graph coloring methodology generalizes the impulse field creation.
  - ■ FEOTS allows for regional simulations from a parent model. Methods are provided to extract the rows and columns of a transport operator for a subdomain of the parent model. This feature is demonstrated in the example simulations presented.
  - ■ The remainder of the offline modeling approach is identical to Bardin 2014 Statements have been added in the introduction to this effect to show how our work is distinguished from Bardin 2014 and other predecessors.

- ● @ 149 I think it makes sense to move this section at the beginning of the paper
  - ○ We have moved this section further up in the paper; reviewer #1 made a similar request.

- ● @ 169 In your notation throughout the paper A is almost always a matrix, which is a linear operator, so (12) would always hold, maybe use a different notation like A(c1 + c2) \neq A(c1) + A(c2) ?
  - ○ We understand the confusion the notation here has caused. The main point we wanted to get across is that the parent model uses a nonlinear advection scheme and the offline model assumes the advection scheme can be written as a linear matrix vector multiplication; in this case, the two are not equivalent. In the presence of smooth tracer fields, the flux-limited Lax Wendroff method is third order accurate and has little diffusion (the leading truncation error is biharmonic). In fields that are not locally monotonic (like our impulse functions), the method reduces to an upwind method that introduces numerical diffusion. We now include discussion that illustrates how the impulse fields result in diagnosing advection operators that are more diffusive than the With requests from the other reviewer to clarify how the flux-limiting results in a transport operator that is more diffusive than Lax-Wendroff (without flux-limiting), we've rewritten these statements about the issues with using nonlinear advection schemes with this

approach. Because the parent model description now comes before the description of the offline model and the operator diagnosis, we've moved this discussion to a new section following the "Graph Coloring to Operator Diagnosis" section.

- @ 200 The FV discretization should be mentioned when you bring up the semi-discrete model
  - We've moved the parent model description before this section and have made changes to define what methods are used to discretize the advection operator. This should clear up this issue.

- @ 203 But is it a physically reasonable choice ? Also, the way you test accuracy is potentially hindered by having a smaller number of significant digits.
  - This is a question that is open for debate and it is not a question this paper aims to address. The focus of this paper is to present to the community a tool for offline tracer simulation and share progress and current capabilities. This example of constant tracer preservation is unique in that numerical errors are zero (proof of this is now included in the paper). Any deviation from a constant value for the tracer is a result of round-off errors accumulating over time. We present a "worst-case scenario" for round-off errors (single precision and aggressive optimization). What we show in these figures is that the error in an $O(1)$ constant tracer field is at worst 3-4 orders of magnitude smaller than the expected answer (a constant field equal to 1). As we indicated to the other reviewer, perhaps the comparison of single to double precision is a topic for another study. Defining physically reasonable comes down to what fields we are trying to measure and with what confidence.

- @ 216-218 Do you consider these to be small errors ? for single precision, machine precision is around 10^{-8} and here it looks like you are several orders of magnitude above that. Can you comment ?
  - The -Ofast flag in gfortran enables a number of unsafe math operations. From the GNU documentation :

    -Ofast
    Disregard strict standards compliance. -Ofast enables all -O3 optimizations. It also enables optimizations that are not valid for all standard-compliant programs. It turns on -ffast-math, -fallow-store-data-races and the Fortran-specific -fstack-arrays, unless -fmax-stack-var-size is specified, and -fno-protect-parens. It turns off -fsemantic-interposition.

    In our experience, Ofast incurs larger round-off errors than machine epsilon for single precision numbers - why this happens may best be answered by a compiler writer.

When using the five day averaged operators, as we are in this case, the transport operators are held fixed over 5 days of simulation time. The volume anomaly grows linearly with time when the operator is held fixed; this includes non-surface layers where the only contributing non-zero values are given by round-off error. Our working idea here is that the round-off errors accumulate over time.

- @ 228 Please add an explanation to describe exactly what you mean by "online"

  We now add the clarification:

  Do the tracer distributions simulated by FEOTS using transport operators diagnosed from the parent model E3SMv0-HiLAT03 (offline) faithfully represent the tracer distributions simulated by the parent model itself (online)?

- @ 236 What do you mean with "inventory"

  We replaced the term *inventory* by *stock* to better express that we are discussing the total amount of dye tracer integrated over a region.

- @ 288 Can you comment on how much the nature of the problem affects the averaging methodology ?
  - We have added the following to the beginning of this section

    The parent model is capable of producing velocity fields that have a wide range of scales of spatial and temporal variability. The shortest temporal periods are on the order of a few time steps and the longest period is the duration of the simulation. In general, higher resolution models introduce more variability on shorter length and time scales and some consideration is needed when selecting an averaging period for the transport operator diagnosis. For storage reasons, it is not practical to store snapshots of the transport operators at every time step. Conversely, representing the ocean transport with long time averages may exclude the effects of important variability. The choice in time averaging period for the transport operators can impact the evolution of tracers calculated in FEOTS and an appropriate balance of practicality and accuracy should be struck.

- @ 290 Then maybe this paragraph should be the first one you discuss in your results section
  - We've rephrased the introduction to this section to read as

    One goal of FEOTS is to perform tracer calculations at a lower computational cost than the parent model. Additionally, FEOTS allows researchers to take advantage of transport operators produced by state-of-the-art climate simulations to conduct regional offline simulations. This provides flexibility in studying ocean

transport phenomena and increases the value of online produced model data while considerably reducing the computational expense for researchers solely interested in studies involving passive tracers. Here we evaluate the computational performance of a regional FEOTS configuration and compare it with the global parent model.

- @ 298 - 299 Can you clarify how the averaging processes works ? For instance, considering your 1 day and 5 day averaged operators, how many runs did you do for each of them ?
  - We've added more details to the Graph Coloring approach to operator diagnosis that illustrates how the impulse response functions are diagnosed and time averaged.

- Table 3 Would it be possible to actually do a run where you report head to head the times for online and offline in seconds to clearly show which one is faster ? Estimated costs do not provide strong enough grounds in my opinion to claim that the benefits of your approach are clear.
  - We have experimented some with running FEOTS in a global configuration and have found that the simulation is fairly slow, although we don't have direct measurements recorded. Additionally, the memory requirement is about 15 GB in single precision (30 GB in double precision). For memory bound algorithms (like finite difference, finite volume, and sparse matrix vector multiplication) this is quite a bit of load to put on a single core. We've rephrased some things in this section to highlight these facts as well as to indicate that we are suggesting that it is probable that a global configuration of FEOTS could provide comparable runtimes to the parent model with fewer CPU-hours. This, to us, suggests viability and motivates us to explore exposing parallelism in FEOTS' core methods.

    We have updated the text in this section to reflect these remarks.

- @ 331 This is why I believe a head to head comparison would be more informative.
  - We had started a few global simulations during this study, but found that the time-to-solution was too high; we simply would not have had enough time to complete this study in the time-frame allotted for this work. This is because FEOTS is in serial; the hotspot profile shows where most of the time is spent while running the code in forward-integration mode. However, our goal in this section was to show the amount of compute resources (number of cores multiplied by the time spent using them) is lower for FEOTS; if this weren't the case, parallelizing would provide no hope of getting the time to solution lower than the parent model and would invalidate our whole motivation for doing this. This particular statement you've highlighted was mentioned to indicate where our focus is going to be in the near future in terms of improving time-to-solution.

- @ 356 Why did you do an estimate of the time instead of actually running offline vs online head to head ?
  - FEOTS is currently written in serial and is considerably slower, at the moment, than running POP with O(1000) MPI Ranks. We simply did not have the time to run such a long serial simulation to go "head to head". However, our estimations suggest that the amount of compute resources required to run FEOTS globally would be lower than the online parent model. This is encouraging when considering whether or not to put effort into parallelizing FEOTS.