

Author response to comments from Referee 2:

We thank the reviewer for their thoughtful review. Referee remarks are shown in red and our responses are given in black font. Changes to the manuscript text are given in blue font.

A comparison of on-line and offline models, the latter using the Transport Matrix Method, is long overdue and will be a welcome addition to the literature, even just as an example for a single model. It's an easy paper to review as the requirements for such a paper are just 2 things: a good description of the method; a sensible choice of parameters to compare. I'm keen to see the paper published but there are a few details that I'd like to see addressed before then...

Method:

- I don't have an issue with the choice of model as the paper is effectively an example and I'm not sure I see the value in a much longer and exhaustive paper doing the same with a variety of models, particularly as they are all evolving. However, I don't see the value of the comparison of how the model performs with and without the FCT scheme.

For the purposes of the paper all that is needed is a base model – it doesn't matter if it performs a little less well than another version. Hence, I would either cut Section 3.1 or move it into an Appendix. As another option, if it is argued that Section 3.1 is there to allow comparison of the offline run to the FCT case then this needs to be done more rigorously by taking the spun-up UW3 model and running it onwards on-line with FCT now turned on for the comparison.

We agree with the reviewer that this comparison detracts from the main purpose of the paper and only included it because it could be of interest to other UVic users who may want to apply the TMM. We have moved Section 3.1 to an Appendix.

- The manuscript is a little vague about the details of the runs for the comparison of the on-line models. The starting point is a 13,000 year spin-up of the on-line UW3 model. The TM is then extracted using an extra year run. Are the offline and on-line models then compared purely on the basis of a single extra year run after the 13,000 year spin-up? I would hope not as unless the off-line perfectly mimics the on-line model there is no guarantee that any transient response of the off-line model will be fast, and the differences may be small purely because the two models have had little time to diverge, particularly if it takes 13,000 years to spin-up. The comparison of the difference between the on-line and off-line models should at least state the time over which the models are run for comparison and this should be at least of order 10 years. My recommendation for minor changes is on the assumption that the models were run for longer than 1 year before comparison. If just one year I think the need to re-run the models for longer would constitute major changes

We're sorry for the confusion. The TMs are extracted from an equilibrium state of the OGCM (obtained after a 13,000 year integration). This only requires running the OGCM for one additional year. The TMs are subsequently used to perform an offline integration of the biogeochemical model for 5000 years to equilibrium. Output from the final year of this run is used in the online/offline comparison. We have added the

following paragraph to the end of Section 2.2:

The offline biogeochemical model is forced with the relevant physical and biogeochemical fields taken from the equilibrated online model. In the present case, these are monthly mean wind speed, insolation, sea ice concentration, temperature, salinity, freshwater flux (evaporation, precipitation and runoff) and iron concentration. All fields, including the previously extracted transport matrices (also at monthly mean resolution), are linearly interpolated to the current time step before being applied. The offline model was integrated with a time step of 8 hours for 5000 years to equilibrium, with monthly averages of various fields from the final year of this run used for comparison with the equilibrated online simulation.

- The manuscript describes the issue with a leapfrog scheme but is a little vague about the compromise made. Is the TM extracted from just one strand of the leapfrog scheme? Is the usual process of blending the 2 strands side-stepped and ignored for the TM?

We apologize for being a bit vague on this point. We have modified the relevant paragraph in Sec. 2.1 as follows to provide additional detail:

A second complication is from the time-stepping scheme, which in UVic ESCM is leapfrog. The explicit horizontal advection and diffusion terms are also sometimes staggered with respect to each other for stability. Both require storing the tracer field at odd and even time steps. While this can be replicated offline, in order to use a common scheme for all ocean models from which TMs have been extracted (e.g., MITgcm variously uses Adams-Bashforth, direct space-time discretization and other schemes), we combine horizontal advection and diffusion into a single explicit transport matrix, A_e which is time-stepped with a simple, forward Euler method. Specifically, to extract the explicit matrix, we only store the (passive) tracer field at the current time step. This field, which is reset to a pattern of 1's and 0's at the beginning of each time step, is then stepped forward by UVic ESCM like any other tracer. The change in the tracer field divided by the time step is the explicit tendency matrix. With this procedure, which does not require changes to the underlying code, the usual leapfrog scheme is side-stepped.

Comparison

- The authors suggest that some of the largest fractional errors come from differences in small values. In Fig. 7 though it looks like there are significant errors associated with large values for diazotrophs. I'd like to see an extra column of plots showing the difference between the on-line and off-line models for these fields as elsewhere in the manuscript.

We have modified the figure (now Figure 2) to additionally show offline-online differences.

- The relative errors in Fig 13 are very noisy. It would help to additionally have the contours for +/-0.5 relative error marked on the panels for on-line and off-line zonally averaged fields.

Figure 13 is now Figure 8. Including contours as suggested made the plot just as

noisy, but changing the format from "shade" to "fill" (these plots are made in Ferret) and adjusting the scale has reduced their noise.

- A couple of minor points but the y-axis labelling on Fig 14 needs an extra decimal place for phosphate and it also seems strange that phosphate diverges (then converges) so quickly from the same starting point if it is just a one year run.

Figure 14 is now Figure 9. The y axis of phosphate has been fixed. Slight differences in the application of external forcing between online and offline models are the likely cause of differences between the simulations, particularly in regions experiencing high seasonality (e.g., the North Atlantic).

We're don't follow the reviewer's comment that "it is just a one year run". Each line on the plot represents an equilibrium solution obtained after a long spin-up integration (13,000 years for the online run and 5000 years for the offline run).