Resolution dependence of interlinked Southern Ocean biases in global coupled HadGEM3 models.

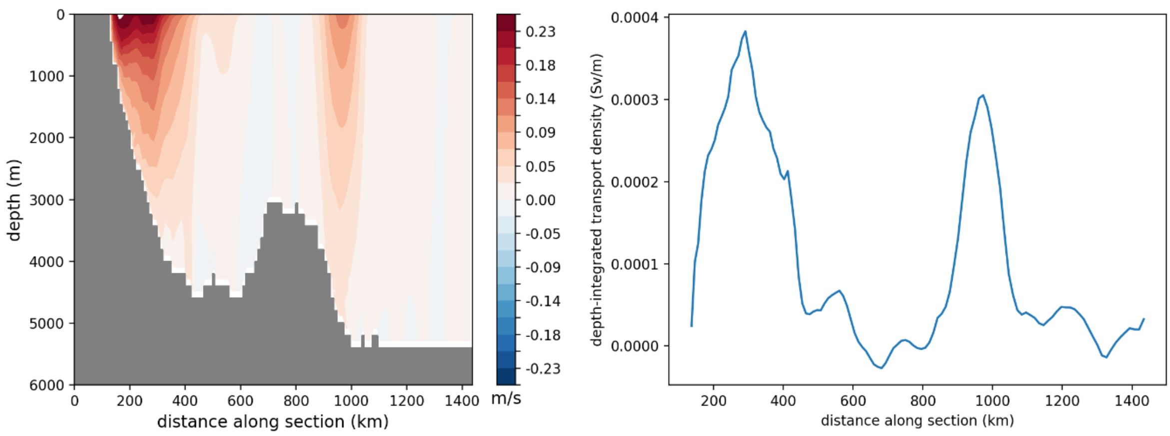
Authors' response to reviewers

We would like to thank both reviewers for taking the time to provide detailed and constructive reviews.

Here we respond to the general comments of both reviewers first and take the more detailed comments later. Where we have not responded explicitly to minor comments and corrections we would propose to follow the reviewer's suggestion.

General comments

1. Separation of the recirculating gyre transport and ASC transport. Reviewer 1 makes the point that the dynamics of the gyre and the slope current are distinct and suggests that we provide figures for the ASC transport separately by calculating the transport over the continental slope. In the region of the gyres it is difficult to separate the ASC from the gyre, since the recirculating gyre transport impinges on the continental slope and models at these resolutions do not form very distinct jets – see for example the plots below showing cross sections of currents and depth-integrated transport densities from the MM model in the Weddell gyre.



In order to give some indication of the circumpolar flow distinct from the recirculating flow we propose to follow the reviewer's suggestion of providing plots of the depth-mean currents in the top 500m to complement the plots of the streamfunction in Fig 1. We will also provide an extra timeseries plot in Fig 6 showing the timeseries of the counterflow at the southern boundary in the Drake Passage (as defined in Fig 2) compared to the observations of Meijers et al (2016).

2. Use of EN4 1950-1954 climatology for assessment. Both reviewers make the point that observations in the Southern Ocean for this period are extremely sparse, and therefore advise caution in the conclusions drawn by comparing the model with the EN4 climatology for this period. The EN4 analysis uses a climatology for the period 1970-2000 as a background climatology to which the solution relaxes in the absence of observations (Good et al, 2013). Given the lack of observations in the Southern Ocean, the 1950-1954 climatology is likely to be very close to this background climatology. For a long-term climatology to compare the model to this is the best we can do, but we will elaborate on this point in the revised text.

3. Length of integration and experimental design. Both reviewers point out that the integrations are short in the context of climate studies, and reviewer 1 questions the statement that *"the early spin up of the model can be useful in diagnosing model biases, since at this stage the model has not drifted too far from initial conditions"* because the initial state of the ocean is out of balance with the greenhouse gas forcing being used and there will be an adjustment associated with this. These experiments follow the HighResMIP protocol (Haarsma et al., 2016), in particular the *spinup-1950* and *control-1950* experiments, in which the model is initialised with "1950s" initial conditions and spun up with 1950s greenhouse gas forcing. As stated by Roberts et al (2019), the nominal spin up for these integrations is very short (30 years) due to computational constraints. As noted in point 2 above the "1950s" initial conditions are likely more similar to a later state of the ocean, so there is an imbalance between initial conditions and the forcing, but the imbalance will not be as great as for a pre-industrial spin up integration. Still it is clear that the model is adjusting in the first decades of the spin up and nowhere near an equilibrated state. However from the timeseries in Fig 6 it seems that there is a fast initial adjustment over the first 2-3 decades after which the three models reach very different states, and these differences in many cases persist for multi-centennial timescales. (We have only shown the first 100 years in Fig 6 because of the length of the sensitivity experiments). In particular, the biases in the MM model appear to be very persistent, so we believe we can learn something about them by studying the initial adjustment period. We will try to make these points more clearly in the revised manuscript.

4. Dense water formation and export. Reviewer 1 suggests some analysis of dense water formation and export from the shelf and AABW export to shed light on the water mass biases on the shelf and possible spurious mixing in the interior. We agree that the models' representation of dense water formation and export could well be relevant to the formation of the biases studied in the paper. From the cross sections in Fig 7 it seems clear that the export of dense water from the

shelf in the Weddell Sea is only captured by the 1/12˚ model. We plan to do a systematic study of the dense water formation in the different models looking at the water mass transformation on the shelf, the export over the shelf break and the time evolution of the reservoir of AABW, but we would prefer to present this in another paper. We propose to include another subsection in the Discussion section, discussing the dense water formation and export and its possible links to the large-scale biases. In this section we would also discuss the possible loss of AABW due to spurious mixing (see point 5 below).

5. <u>Explanation for isopycnal slumping in the open ocean</u>. Reviewer 2 says that the slumping of isopycnals across the ACC in the medium resolution model is counter intuitive and that this is not addressed very prominently in the paper, with a tentative explanation only appearing in the last paragraph of the Conclusions. We agree and propose to follow her suggestion of emphasising the counter-intuitive nature of the result in the Discussion section with a discussion of the possible link to spurious mixing and the loss of AABW in the new subsection on the dense water formation and export.

6. <u>Inclusion of partial slip and scale-aware Gent-McWilliams in 1/12˚ model</u>. Reviewer 2 asks if we tested partial slip or scale-aware Gent-McWilliams in the 1/12˚ model. We have included the combination of partial slip and scale-aware GM in our standard 1/12˚ model configuration, but due to computational expense we only have a clean test of the impact in a forced ocean-ice configuration. In this test, the combined effect of partial slip and scale-aware GM is still positive in the sense that it increases the ACC strength and reduces the gyres, but the impact is not as great as it is for the 1/4˚ model. This might be expected because the increased resolution at 1/12˚ means that the GM coefficient will be non-zero over a smaller area than for the 1/4˚

<u>Detailed comments reviewer 1</u>

L16: "Antarctic Overturning Circulation". I think stating "between the near surface and deep ocean via the formation of Dense Shelf Water (DSW), Antarctic Bottom Water (AABW), and mid-depth ocean via mode and intermediate water formation" would work better here. If you are linking to the Southern Ocean being critical to the climate system - also mentioning the mode & int water formation is important as this is where all the heat and carbon is going. We agree that an expanded description of the overturning circulation would be good here.

L24: Also, there are a lot of warm biases in 1-degree CMIP-class models (Beadling et al., 2020), not sure there is a definitive link to high resolution models being warmer? For example, in some high-res simulations we actually see warm biases improve with finer nominal grid spacing. Agreed – the statement about the resolution dependence may have been a Met Office centric point of view. We will remove that and just make the statement that warm biases in the Southern Ocean are a common problem in CMIP models.

L67: “Cavities under ice shelves are closed and the output of basal melt water at the ice shelf front parametrised as described in Mathiot et al. (2017).” This is interesting, so this model represents “ice shelf melt”? Is this just based on some threshold of solid precip over the Antarctic continent? Could you elaborate on this? I assume this does not imply there are realistic melt rates? The distribution of melt water input around the continent is based on Rignot et al (2013), but for these experiments the overall magnitude of melt water plus iceberg calving is scaled to equal the total precipitation over the Antarctica at each timestep, ie. An assumption that the total mass of ice over the continent is constant. We will add a sentence to this effect.

L84-85: “However the early spin up of the model can be useful in diagnosing model biases, since at this stage the model has not drifted too far from initial conditions.” ß I am not sure I agree with this statement, assuming the ocean is starting from a present-day climatology (say WOA13 or WOA18) and a pre-industrial atmosphere, this early stage is an unrealistic climate state and an assessment of realism (i.e., biases relative to observed) is better made once the model has been able to achieve its own equilibrium (or better yet,reached that equilibrium and forced with observed climate forcings; i.e., the historical simulation). I tend to think of the spin-up stage as the adjustment stage that we don’t want to consider when doing assessments against observations. I suggest rewording this or expanding on your reasoning here. See response under point 3 above. We will rewrite these sentences to make our approach clearer.

L93-101: As you note, the gyres and ASC transports merge into one another particularly in the Weddell, so it is hard to discern these from one another in the current figures. I would suggest an additional plot of the upper 1000 m speed (or upper 500 m speed) to see the differences in the strength and location of the ASC. L93-101: As you note, the gyres and ASC transports merge into one another particularly in the Weddell, so it is hard to discern these from one another in the current figures. I would suggest an additional plot of the upper 1000 m speed (or upper 500 m speed) to see the differences in the strength and location of the ASC. As discussed under point 1 above we will follow this suggestion and provide maps of the depth-integrated currents over the top 500m to complement the plots of the the streamfunction in Fig 1.

L102: “net eastward transport” this wording is confusing, there is eastward and westward flow through Drake Passage, should this just say “net” to avoid confusion? We will use “net transport” rather than “net eastward transport”.

L103-104: It is worth mentioning that this is exceptionally weak even compared to earlier / other estimates (Cunningham et al., 2003; Griesel et al., 2012; Meijers et al., 2012; Koenig et al., 2014; Firing et al., 2011; Xu et al., 2020). We will add this point.

L109-111: It is worth mentioning that Xu et al., (2020) also shows net westward flow at depth - this is why the authors argue that Donohue et al., (2016) overestimated the net transport through Drake Passage. Although they are referring to bottom recirculations - different from what is shown here. We will add this point.

Figure 3 caption: Why is the 1950-54 climatology used here for comparison? Because it is close to the initial conditions? Yes. Although it is also the case that the 1950-1954 climatology is likely very similar to a later climatology (see point 2 above), so it could be viewed as comparing to a "best available" long-term climatology.

L122-128: How different are the sea ice edge locations? It would be helpful to add these to the plots of Figure 5 for reference of where these anomalies are. We will add lines showing the maximum sea ice extent.

L131-133: Given that the gyres and ASC are governed by different dynamics, I strongly suggest breaking this down into an assessment of gyre strength and the ASC separately. Addressed under point 1 above.

L143: "and comparing to a similar average performed on the 1950-1954 climatology of the EN4.1.1.g10 analysis dataset". I might be missing something, but why is this time period used for comparison? Observations would be very sparse for this time period and particularly so in the Southern Ocean. Addressed under point 2 above.

L157-158: Yes, this could indicate and issue with CDW cross-shelf intrusions, but this could also be due to the westward transport of cold, fresh Weddell Sea water around the WAP mixing with CDW (making it colder & fresher). The maps of ocean velocities support this connection. This has been found in other simulations when the ASC accelerates (Beadling et al., 2022) and this mechanism has been documented as well by Morrison et al., (2023) "Weddell Sea Controls of Ocean Temperature Variability on the Western Antarctic Peninsula". You mention this below in lines 168 – 169 but it should be mentioned here or this discussion combined. We will add something about the advection of fresh water at this point.

L163-165: This would be shown nicely with a surface water mass transformation analysis (sWMT). This is also consistent with Tesdal et al., (2023) which showed that when the ASC accelerates, DSW reduces as the shelf becomes more buoyant. We agree and we plan to look at water mass transformation metrics as part of future analysis (see point 4 above).

Figures 1,3,4,5,8. It would make comparisons easier for the reader to add two additional panels to these plots of the N216-ORCA025-GM MINUS N216-ORCA025 and N216-

ORCA025-PS MINUS N216-ORCA025. It is hard as of now to discern some of the differences. We are a bit reluctant add more panels to the existing figures as there are a lot of plots already and we think that (perhaps with the exception of Fig 2) the differences between the N216-ORCA025 experiment and the two sensitivity experiments are fairly clear by eye (eg. Weddell gyre strength in Fig 1 or Amundsen Sea temperature in Fig 4). We propose to provide difference plots but to include them as supplementary figures as suggested by reviewer 2.

Figure 2 & L200-203. It looks like GM really impacts the westward along-slope flow (ASC) through the passage while the rest appears unchanged. PS appears to reduce flow everywhere (even the eastward flow in the Subantarctic Front), however reduces the eastward components more … which is why the net increases. It is hard to see visually what component of the along-slope flow is decreasing --- is this mostly coming from the bottom flow or surface intensified flow? We agree that it is hard to see the details of the impact of PS and GM in this plot – this should be clearer in the supplementary difference plot.

Figure 6: The lines for N216-eORCA025-GM and N216-eORCA025-PS are very hard to discern. Can you make one have circle markers? The dashed and the dashed-dotted are very hard to distinguish. We will experiment with alternatives to make this clearer.

L205: "The timeseries show that again, Gent-McWilliams appears to have a stronger impact than partial slip." This sentence is referring to shelf salinity, yet this is only true for the Ross. The PS West Antarctic shelve Amundsen / Bellinghausen) looks better for the PS (Figure 3) (This is ALSO true for shelf temperature as you mention below, so this would just require rewording). The timeseries for the Weddell salinity looks similar between the two. We will clarify this in the text.

All figures: Increase size of text on color bars / axes, some of these are hard to see. We will adjust these.

Figure 7:

- The top figures from Thompson et al. (2018) have a y-axis in km, but the rest are in m. This should be made to be consistent across the panels. Text on axes are also hard to read. The titles on the top and bottom also look very large compared to the other labels in the other figures in the manuscript. We will follow these suggestions.

- Figure 7: I assume that the grey shaded regions are not the models true bathymetry? The blocky-nature makes it appear that the models do not account for partial cells. I assume in reality this is more smooth? It is true that the masking in the plots does not include partial cells, and that if this were done, the bathymetry would look smoother. However, partial cells have the biggest smoothing impact in weakly sloping bathymetry. Over the shelf slope the bathymetry is still quite blocky even with partial cells.

• Figure 7: The Thompson et al., figures are conservative temperature, not potential temperature. These should be consistent between the observation panels and the model output panel. Thanks for pointing out this oversight – we will correct it.

L276: This is consistent with the feedback to meltwater Bronselaer et al., (2018) suggested i.e., slumping of isopycnals resulting in more heat delivery to shelf --- just pointing it out, but perhaps not necessary to discuss. Yes that's a good point. The mechanism described in their paper is similar to the one we are proposing here.

Detailed comments reviewer 2

L. 30-32: It would be appropriate to also mention the emergence of models with unstructured grid configurations, as they are an approach to overcoming the issue with affording highenough resolu:on to resolve high-la:tude eddies in global models, e.g. FESOM (Wang et al., 2013; Scholz et al., 2019), ICON (Jungclaus et al., 2022;Korn et al., 2022). We will add something about unstructured models.

L. 61-62: "A free-slip boundary condition [...]" - It would be valuable to also mention what viscosity scheme is used, and potentially also the parameter settings since they likely differ between resolutions, as these also have the potential to affect the flow and thus the biases. Agreed – we will add the information about the viscosity choices.

L. 64-65.: "Diffusion of tracers along isopycnal surfaces, parameterising eddy mixing [...]" – Is there any regional reduction of the parameterised eddy mixing around the equator where eddies are fully resolved, in particular in the higher resolutions? No. There is a reduction of the diffusion coefficient with reduced grid spacing at higher latitudes to avoid numerical instability.

L. 75: "with the N216 atmosphere" - Are the results consistent or at least similar if one of the other atmospheric resolutions are chosen? This would indicate that the results are more widely applicable to other coupled models, regardless of what atmospheric setup they use. We have not looked at the Southern Ocean biases in the other HighResMIP integrations in detail, but Roberts et al (2019) show that for the ACC transport, a change in atmosphere resolution makes little difference to the long-term behaviour (their Fig 18). For the SST biases there is some impact of atmosphere resolution (their Fig 7) as might be expected.

Page 3, footnote 1: Was the eddy-permitting model ever tested with no-slip? It would be useful to motivate the choice of partial-slip over no-slip in this case, and discuss how choosing a no-slip condition might have impacted the overly strong ASC. Useful references may be Penduff et al., 2007; Deremble et al., 2011; Nasser et al, 2023. The choice of free slip for the 1/4° model goes back to Barnier et al (2006) and Penduff et al (2007). We have not tested the 1/4° (or 1/12°) model with no slip, although with hindsight that might have been an informative thing to do. Any choice of lateral slip condition in z-level models is difficult to justify on a physical basis, and we tend to see this result as an indication that the large scale biases we are looking at may be linked to poor representation of bathymetry-flow interaction in the model.

L. 93-95: "As well as having more active gyres, the higher resolution models also have a stronger ASC […]" - In the next paragraph, you mention the consequences of this feature in the Drake Passage, but only after you discuss the over-flattened isopycnals. This makes this part of the text feel somewhat fractured. Maybe mention the Drake Passage briefly already here, or rearrange the next paragraph to discuss the ASC behaviour before the overly flat isopycnal slopes. Following the suggestion of reviewer 1 we intend to include an extra figure showing the depth-mean currents for the top 500m. This should illustrate the circumpolar nature of the ASC in the eddying models – we will mention the westward flow in the Drake Passage when describing this figure.

L. 98-99: Klatt et al. (2005) is one of few observationally-based estimates of the Weddell Gyre strength published in recent decades. It is, however, not the only one. Reeve et al. (2019) estimate it to 32 +/- 5 Sv based on ARGO data. Older observational estimates by Farbach et al. (1991) and Yaremchuk et al. (1998) are also lower c.f. Klatt et al. Meanwhile, the transport in models ranges between at least 10-80 Sv (see e.g., Neme et al., 2021; Wang, 2013). We will mention the other observational estimates in the revised text.

L. 105-106: "The weaker ACC transport in the higher resolution models is associated with a flattening of the time-mean isopycnal slopes in the Drake Passage" - Given that particularly the ORCA025 resolution is not fully eddy-resolving in the ACC region, it is somewhat counterintuitive that there is an over-flattening of the isopycnals, which suggests too much eddy compensation. From reading the text here, it makes one wonder how that can be. As mentioned in the general comments, this counterintuitive behaviour is mentioned by the authors later in the manuscript, but should be acknowledged sooner. However, looking at the figures, it looks like the isopycnals are very steep (steeper than ORCA1) in the northern part of DP, where the core of the ACC is, and then flattened in the centre, where there is weaker (ORCA12)/counter(ORCA025) flow. In ORCA12, steeper isopycnals than in ORCA1 are also observed between -63 and -62 degN. These details should be mentioned, as they might help elucidate what is actually happening to the ACC transport. For general response see point 5 above. We will include a more detailed description of the isopycnal slopes in the Drake Passage in the revised manuscript.

L. 107: "counterflowing currents […] associated with the Shackelton fracture zone" – The authors mention that the modelled counterflow along the southern shelf break is unrealistic c.f. Meijers et al. (2016), but they give no indication of whether counter flows at the Shackelton fracture zone correspond to observations or not. Xu et al (2020) show complex recirculations in the Drake Passage, especially at depth, in their 1/12° model and argue that these recirculations, if realistic, are challenging to sample with observational arrays. We will make this point in the revised text.

L. 122-128: On resolution-dependent temperature biases - What are the differences in global mean surface temperature in these simulations, and compared to the observational dataset? As some of the biases in SST can stem from the atmospheric model, or results of the difference in resolution in other regions, it might be useful to make a supplementary figure where the biases are normalised to the observational global mean surface temperature. The HadGEM3 HighResMIP models tend to have cold SST biases away from the Southern Ocean and these cold biases tend to reduce as the ocean resolution is increased (Roberts et al, 2019, Fig 7) mainly due to improved representation of boundary currents and ocean heat transports. We agree that the attribution of SST biases is more complex than for some of the other biases examined in the paper because of the direct influence of the atmosphere and possible teleconnections, but given the complexity of the global picture we aren't sure that a comparison with the global mean SST bias will be very informative here.

L. 165: "in the eddy-permitting model [...] deep water formation has been suppressed" - Does this lead to (more) unrealistic open-ocean convection than in the other two model versions. In the control runs, both of the eddying models develop regular open-ocean convection after a few decades. The 1/4° model seems to be no worse than the 1/12° model in this respect.

L. 185-189: On introduction of the scale-aware G&M: What are the implications of running a resolution that to some degree allows eddy formation, but then also parametrising eddies on top of it using G&M, which can lead to "smoothing out" of the actual eddies? This should be clarified in the text. The idea is similar to that in Hallberg (2013) where we only "switch on" GM where the model fails to resolve eddies, so that we don't smooth out the eddies where the model is eddy-resolving. But of course this is difficult to do in a precise way and the whole question of how to parametrise eddies in partially-eddying models is a big research topic as we have noted in Appendix B. We will expand on this a bit in the main text.

L. 190-192: On the introduction of the partial-slip condition – As mentioned above, it would be useful to motivate why the choice was to go with partial-slip and not no-slip for this resolution. Also, based on the description of how partial slip affects the biases in ORCA025, it seems that introducing it south of 50S in ORCA12 could potentially fix the remaining biases with counter flows in the Drake Passage, and thus with cold waters being advected around the Antarctic Peninsula in this resolution as well. As discussed under point 6 above we have included the scale-aware GM and partial slip in the Southern Ocean in our standard 1/12° model and the combination of these two changes does appear to reduce the Drake Passage counterflow in this model.

L. 223-224: “the Fresh Shelf, the Dense Shelf and the Warm Shelf. In Figure 7” - It would help guide the reader’s eye if it were also indicated in the figure and/or the figure caption what column exemplifies which one of these three regimes. We agree this would be clearer and will add these labels to the figure.

L. 238-239: About the V-shaped pattern of isopycnals in the higher-resolution models - I struggle to find this V-shape in the drawn isopycnals in the eddy-permittng resolution. If so, I can only see it in the two isopycnals labelled 27.5 (this might be a mistake in the labelling, as it occurs twice). We agree that the eddy-permitting model only has a shadow of the V-shape – and it also clearly fails to represent the cascade of dense water. We will update the text to make this clearer.

L. 262-265: The EN4. 1 climatology does not show the same steep ispoycnal slopes near the continent as the model in this region but, as mentioned, the observations included in the climatology are sparse. It could be useful to cite other data (not included in the EN4.1 climatology) that give indications about the isopycnal structure in the area even if those are from from a later time period. Pina-Moleno et al (2016) show observations with a strong ASC and associated front in the same sector of Antarctica. We will cite this reference here.

L. 310-320: On open-ocean polynyas - Are these events stronger/more frequent in the eddy-permittng c.f. the eddy-rich resolution, given that the latter appears to have some more capability of forming dense water on the shelves? (see also L. 165) See response to L. 165 comment.

L. 332-336: About biases along the continental slope/shelf – Here, it would be helpful to clarify which factor is the more important in reducing these biases: adding G&M or introducing the partial-slip condition. Based on the timeseries in Fig 6, they seem to be equally effective for most bias metrics, but GM seems to have a stronger impact on the Ross gyre strength and the Western Ross salinity. We will add a sentence to that effect.

Figure 5: mean SST-lines - It is unclear to me which mean SST this refers to, and as the lines are completely unlabelled, there is no indication of what temperatures the different lines represent. The lines in Fig 5 are mean SSH and are simply there to help locate the SST biases with respect to the gyres and the ACC.

Figure 9: In this figure, it might be more illustrative to show the model results as anomalies from the reanalysis data. In the other figures, observational datasets are shown in the bottom-right subpanel. It would be helpful to keep the same structure throughout the manuscript. We would prefer to keep these as plots of the plain fields rather than anomalies, in line with the approach in the other figures, but we will provide anomaly plots as a supplementary figure. We will move the JRA plot to the bottom right as suggested.

Figures overall: It could be helpful with supplementary figures showing the differences between the standard N216-ORCA025 and the other two ORCA025 versions (GS and PM) as anomalies from the standard (GS-standard, and PM-standard). We will provide the difference plots as supplementary figures. (This was also requested by reviewer 1).

References

Barnier et al (2006): https://doi.org/10.1007/s10236-006-0082-1

Good et al (2013): https://doi.org/10.1002/2013JC009067

Hallberg (2013): https://doi.org/10.1016/j.ocemod.2013.08.007

Pina-Moleno et al (2016): https://doi.org/10.1002/2015JC011594

Rignot et al (2013): https://doi.org/10.1126/science.1235798

Roberts et al (2019) : https://doi.org/10.5194/gmd-12-4999-2019

Penduff et al (2007): https://doi.org/10.5194/os-3-509-2007

Xu et al (2020): https://doi.org/10.1029/2020JC016365