Review Trossman et al.

State estimates like ECCO take an important place in our field. It bridges a gap between observations and numerical models. Like observations and numerical models, it has its own pros and cons. This paper particularly looks at vertical mixing resulting from the state estimate and tries to understand and improve this number. It does so by first establishing how good it is (comparing to observational products), and then trying to add the tracer oxygen, to see if this could provide extra information to constrain and improve vertical mixing estimates in ECCO. Although it seems to do so, the results are somewhat disappointing, as the authors also notice. Regardless, this is a result in its own and therefore worthy contribution to publish. Before publication, I do think there are aspects of the analyses that need to be addressed and places where clarification to be made. Hence, with some major revisions, I think this could become a good publication.

Comments are in order of text and thoughts. Major comments indicated.

L20-25 – GM is not mixing, I think. It is advection.
L35 – I don’t understand what is meant with the rotation derived component... Is this about along isopycnal rotation? Perhaps better to explain this more carefully.

**Major L115** – The results are compared to observations of mixing. This is good. But I think it would be worth to compare the results against the parametrization of Lavergne et al 2020. This is semi-observational estimate compares well to Whalen’s work, but covers all grids and full depth. Without data-gaps this may provide better comparison product for the model. The work is also “averaged”, like that from ECCO, and not instantaneous like that from observations. Another reason why this might be a better comparison for this effort. Of course, neither observation, nor Lavargne’s method are perfect for comparison, but having them both really adds value.

**Major - L120-125** - I think you should update WOA13 to WOA18, at least 5 years more data, including BGC data. This is certainly worth the upgrade, and it should be straightforward. Also, which is used, monthly means, annual means?

L170 – for by?

**Major - L200** – N² is used from WOA13. However, WOA13 has problems of its own: to start with is is averaged on isobars and not on isopycnals. This leads to false mixing and we also see that N² is often unstable. However, on average time scales of months, the ocean is “stable” as overturns generally have much shorter duration. Have you made N² stable? There are simple tools for this in the TEOS-10 manual. I think this is important for your comparison.
Major - L225-230 – Equation 1 and surrounding text. The results are sensitive to prior choices made by the modeler, in the form of weights. Here a certain choice is made, but sensitivity to this choice is not studied. The sensitivity is both to the equation weighting (as done here), and to the variable weighting. The latter means that the a-priori estimate of the variable, influences the result. This should as well be considered somehow because this a-priori estimate can be wrong or vary within a range. I see no sensitivity to either of these choices. Or at the very least, a discussion on this possibility. Places to read about this are McIntosh and Rintoul 2001, but also Groeskamp et al 2014.

I’m aware that this can be a pain and sometimes limited by computer power. I don’t know the situation for this study. I therefore strongly recommend to do it if you can, it will strengthen the confidence in your model and the results. IF you really can’t do this, then at least discuss it.

240-245 – Below eq 2, do you here mean to fill in equation 1 into 2, so you then get a “y”. You refer to “y”, where there is none in eq 2.

Section 2: You focus on analyzing k_rho. But what about all other variables. It could well be that if you better constrain k_rho, error is increased on other variables. Hence, how do you keep watch that the rest of the variables are not underperforming?

Line 283-284 – it says, “not shown”. I would have liked to see this in the same figure, I think. This would be informative, and it sounds like you have the data anyway.
Major - Section 3.1 – Observation are instantaneous profiles in a time varying field. We here are dealing with averages in a model. So I was wondering, is taking an average to compare the right thing to do? Is it not better to check individual profiles and add their error together somehow? I mean, averaging a diffusivity is always a strange thig to do. Oké, we must do something, I agree with that. But it seems like many profiles that are physically unrelated are now averaged in one big profile that therefore loses all sense of information. Is there not a possibility to compare and show some selected individual profiles and come up with a metric that addresses the error between the obs and the model, for the sum of each individual profile? I think that would be more meaningful.

Grid lines in Fig 3. And Fig. 7 would be very helpful.

L295 – Do air-sea fluxes influence the results here or is this stuff too deep. What about advective processes?

L310 - There is no table 2.2.

Section 3.3 discussion figure 5. How would this figure look if the differences where small? I mean, it would still be red or blue right? It only gives a sign, but does it say something about the magnitude of the mismatch? It only seems to give a direction of the mismatch. Is it not better to also say something about the magnitude of the mismatch?
L342 – We cannot compare white and white. Please make these locations referred to, gray in figure 6.

Major - L360 – Oke, $O_2$ gives extra information. But which part? Is it the vertical gradients of $O_2$? If so, when do they give extra information compared to $N_2$? It should be possible to argue this upfront and discuss (not do, yet) which other tracers could be of interest in future work. That is, when $N_2$ does not have vertical gradients, but $O_2$ does, then for that region it is worth adding $O_2$. If both don’t have gradients there, then probably some other tracer may work. A discussion on this would be helpful. Possibly a figure.

Section 4 – Could oxygen also increase the errors? It is not said that things get better. I think it is argued that the observation of $O_2$ is accurate and thus it provides information. But for some places where there is lack of data or the values are not averaged the right way in constructing the climatology, errors could be substantial. You have regions where estimates improve, and where they get worse. This could be why. This is related to the previous comment. Again, a discussion would be helpful.

L380-381 – I think with everything going on in ECCO also leading to errors compared to the real world (e.g., parameter choices), I find it is too strong a statement to simply say that data alone is insufficient. You do tone down this point a little in the sentence below, but I still think this statement as written down here is too strong.

As mentioned before, I think this might also be a place to discuss that the data product itself (WOA) has errors in there due to interpolations, and averaging techniques (e.g., horizontal instead of isopycnal averaging).

- C. de Lavergne et al (2020)
A parameterization of local and remote tidal mixing

Journal of Advances in Modeling Earth Systems 12 (5), e2020MS002065


Do Box Inverse Models Work? Peter C. McIntosh¹ and Stephen R. Rintoul¹ 1997

DOI: https://doi.org/10.1175/1520-0485(1997)027<0291:DBIMW>2.0.CO;2