We thank the reviewer for his latest reading of our manuscript, and for providing comments below (in black). Our responses appear in blue font.

Review of:
The interpretation of temperature and salinity variables in numerical ocean model output, and the calculation of heat fluxes and heat content
By McDougall, Barker, Holmes, Pawlowicz, Griffies and Durack

Summary and recommendation

This main goal of this paper is to develop arguments seemingly making it possible for the potential temperature (θ) of EOS-80 based models to be `interpretable' in some sense as Conservative Temperature (Θ), which if true, would allow ocean modellers to directly compare the θ of EOS-80 ocean models with the Θ of TEOS-10 ocean models, instead of comparing the like for like, which has been the accepted practice so far.

This is not quite what we have done in this manuscript. The first part of this sentence is fine, but from "which if true ...", should be replaced with "which if true, would allow ocean modellers to directly compare the temperature output of EOS-80 ocean models with the temperature output of TEOS-10 ocean models."

This paper is difficult to review and understand because it relies nearly exclusively on nonstandard arguments and abstract reasoning, as well as on nonconventional views about the nature of ocean models and of their dependent variables. In my first review, my initial reaction was that the paper had to be wrong, but I was not able to fully pinpoint exactly why. Having now read the paper 3 times and having had more time to reflect on its message, I now understand that the primary cause of my discomfort is the fact that this paper appears to assume that the evolution equation and boundary conditions satisfied by a physical quantity have some bearing on the definition of such a physical quantity, which seems to conflict with what is normally assumed in standard physics (at least, the way I understand it).

We do indeed say that the interpretation of an output variable of a model depends on the evolution equation and the boundary conditions of that variable in the model. Along with its initial condition, the evolution equation and boundary conditions fully determine the values that a variable takes throughout its evolution.

Take potential temperature for instance. As is well known, such a physical quantity is generally regarded as being rigidly defined as the notional temperature that a parcel would reach if brought adiabatically to the ocean surface at the mean atmospheric pressure which is sufficient to fully define it. In particular, the evolution equation and boundary conditions that we may then formulate to predict its temporal definition are not normally supposed to have any bearing on its definition. Indeed, according to my understanding of physics, the definition of a physical quantity and its assumed evolution equation are generally regarded as entirely separate businesses. It follows that if we decide to approximate the evolution equation and boundary conditions for θ instead of using the most accurate one available, the usual view is that this will only introduce errors and uncertainties in the simulated θ, but not alter the character of θ itself. In this paper, however, the authors appear to take a different view. Specifically, they contend that if the evolution equation used to predict the temporal behavior of θ is not the most accurate one available, but an approximated version of it that resembles the evolution equation for Θ, then θ loses its θ character somehow to assume that of Θ. Thus, even if θ is initialised with observed values of θ and if the equation of state used assumes θ as its argument, the authors contend that θ will drift towards Θ after some `long spin-up time' if θ is treated as strictly conservative. (The authors do not provide the evolution equation supposed to be satisfied by the drift, nor do they discuss the physical quantities controlling the relaxation time scales controlling the drift, which would allow us to check the authors' views).

This discussion does indeed reflect what we have said in the manuscript. The air-sea heat flux, and the lack of interior non-conservative source terms are both consistent with our initial hypothesis that the temperature output of an EOS-80 ocean model can be interpreted as being Conservative Temperature. With this as a possible interpretation, we then investigated the implications of making this interpretation, given the different equation of state that the EOS-80 model uses. We concluded that this hypothesis is a good one, and importantly, it avoids an embarrassing error in the air-sea heat flux that is associated with the usual potential temperature interpretation. While we have lived with this embarrassing error for more than a century, it is good to finally be rid of it.

Having realized that the main cause of my discomfort was due to the authors allowing the evolution equation and boundary conditions satisfied by a physical quantity to interfere with its definition, in contrast to what I think is the normal practice, it became much easier to understand the reasons for otherwise many very unclear statements and assertions. For instance, it now made more sense to me why Prof. McDougall and the authors would contend that potential enthalpy should be regarded as the variable defining heat content in the ocean. Indeed, a review of the literature on the subject (Bryan 1962; Bacon and Fofonoff 1996; Saunders 1995; Warren 1999) prior to McDougall (2003) clearly reveals that the quantity $c_{p0}\theta$ used so far as the definition of heat content had been regarded as some approximation to the non-mechanical part of the total energy, its 'heat' quality resulting from the difficulty to convert it into mechanical energy, as per the second law of thermodynamics. Indeed, what standard thermodynamics tells us is that the 'heat' forms of energy cannot be converted with 100% efficiency into 'work' forms of energy. In the classical view, therefore, 'heat' is regarded as a property of the fluid, as a form of energy that is difficult to convert into mechanical energy, irrespective of how surface heat transfer affects it. The authors appear to take a completely opposite view, however, namely that potential enthalpy is the relevant definition of heat because its surface flux captures the entirety of the surface heat transfer, irrespective of its degree of convertibility with mechanical energy. This is why in my first review I expressed the opinion that Prof. McDougall's approach was idiosyncratic, not to cause offence, but to point out how radically different its premise appeared to be compared to that of previous approaches. My criticism was addressed to the fact that the authors appeared to present their arguments in supportof potential enthalpy as the relevant definition of heat as being self-evident, without mentioning to the reader how different its premise is compared to that of previous approaches, nor its ad-hoc character. For instance, adopting the authors' views, how would one define heat if the ocean were in fact primarily thermally forced along its uneven topography? As potential enthalpy referenced to a spatially varying bottom pressure? But then, heat would be a function of potential temperature, salinity, and horizontal position. Would that be acceptable? From a more fundamental viewpoint, shouldn't one seek a definition of heat that is equally applicable to the atmosphere as to the ocean?

It is not that the boundary conditions and the evolution equation "interfere(s) with the definition" of a model's temperature variable. Rather than interfering with a definition, what we have done is to offer an explanation of a model variable. We make the point that if one chooses to interpret the temperature variable in an EOS-80 model as being potential temperature, then this comes at the cost of a substantial error in the air-sea heat flux; the ocean receives more heat than the atmosphere delivers to it! The alternative interpretation, namely that the EOS-80 model's temperature is Conservative Temperature does not suffer from this problem. Rather it suffers from the problem that the horizontal density gradient is a little in error. It is indeed disappointing that not all ocean model codes have converted to the ten-year old equation of state, TEOS-10, but nevertheless, this manuscript will help those analyzing and comparing the outputs of ocean models based on EOS-80 and TEOS-10.

Regarding the remainder of the above paragraph, the non-conservation of Conservative Temperature has been extensively discussed in the literature in McDougall (2003) and in Graham and McDougall (2013). These papers discuss how the geothermal heat flux at depth can be converted into a flux of Conservative Temperature.

To go full circle, the authors also contend that the salinity variables used in models should be interpreted as preformed salinity, on the grounds that current ocean models treat such variables as strictly conservative, which the authors argue is true only of preformed salinity. According to the authors, both temperature and salinity variables will 'drift' towards Θ and S^* after some undefined 'long spin-up time' regardless of how they are initially defined or initialized, provided that they are treated as strictly conservative.

Yes, that's right. Preformed Salinity is a conservative variable and is almost exactly equal to both Absolute Salinity and Reference Salinity at the sea surface (where the sources of freshwater in the model are located). The CMIP models are typically spun up over some 2000 years which is plenty of time for some random initial condition to be flushed out of the coupled system.

Because my understanding of physics is that the assumed evolution equation and boundary conditions of a physical quantity have no bearing on the definition of a physical quantity, my view is that the paper is based on unsound physics. Now, I also must acknowledge the fact that this paper touches on fundamental aspects of physics that are rarely if ever explicitly discussed. The fact that such eminent oceanographers appear to have such a different understanding of physics than I, combined with the fact that the views expressed in this paper did not appear to bother the second reviewer, Prof. Fox-Kemper, a lead author of a chapter in the latest IPCC report, suggests that the issues touched upon are not well understood by the community, and hence that there may be value in publishing this paper along with my review in order for the community to reflect on where it stands on the issues discussed.

Rather than interfering with a definition of a variable, what we have done is to offer an explanation of the model's variables. When a coupled model is run, there is no high authority who decrees what the model variables are. Rather, these models contain evolution equations and boundary conditions, which together, dictate how all the variables evolve. It is then up to we scientists to interpret the variables in a consistent fashion that is cognizant of the equations and boundary conditions to which the variable has been subject during the running of the model.

Major issues

Potential source of divisions – In the event that only part of the ocean modelling community adopts the authors' recommendations, with the remaining part disagreeing with them and therefore sticking to the currently accepted practices, what would be the authors'suggestion for resolving the resulting schism in the community? Wouldn't it be wise/useful for the SCW [the JCS?] or the CLIVAR ocean modelling working group to organize some kind of world-wide poll about the issues discussed to identify to what extent oceanographers agree or disagree with the authors' view that it is ok for the evolution equation and boundary conditions to interfere with the definition of a physical quantity? I hope that the authors can agree that the development of incompatible ocean modelling practices cannot really be good for the field and is likely to complicate the writing up of the 'ocean' chapter of the next IPCC report. It seems to me that the authors should address this issue in their paper, i.e., thepossibility that not everyone will agree with their recommendations.

As with the development of TEOS-10, our view is that the science should first be published, and then the community can absorb what is published, and a consensus will likely emerge.

Specific comments

Line 68 – Saying that density depends on 'heat content' is dangerous and provocative since the concept of 'heat content' is controversial and likely to remain so for the foreseeable future. Why not stick to uncontroversial and non-provocative practices?

Thanks. We have adopted this suggestion; "temperature" now replaces "heat content".

Lines 70-71 – I don't understand the point here. The evolution equations for potential temperature and conservative temperature are non-conservative, whether we think it is useful or not. Treating such quantities as conservative necessarily entails an approximation that is the modellers' decision and has nothing to do with the conservativeness of the numerical schemes.

Here we discuss what ocean models actually do; they treat their temperature variable as being a conservative variable, so we have made no change to the text.

Lines 73-74 – Some numerical ocean models formulate their temperature equation in advective form, in which case the said conservative property is not satisfied (as far as I am aware).

All ocean models known to us formulate their temperature evolution equation in flux-divergence form, without any non-conservative interior source terms. This common practice was documented in Section 9.1 of the review paper Griffies et al (2000) (Ocean Modelling, vol 2, pages 123-192) where it is stated that "All models employ flux form advection schemes, hence allowing for trivial conservation of total tracer amount." That paper documented the practice in 12 community ocean models circa 2000. In the subsequent 21 years, there has been no change in regards to the practice of how ocean models formulate their tracer transport equation. Although numerical methods differ, the underlying equation remains flux-form.

We observe that some atmospheric weather models make use of the advective form to facilitate their use of semi-Lagrangian numerical methods. But it is well documented (e.g., Lin and Rood 1996, MWR vol 124; Lin 2004, MWR vol 132) that such advective form semi-Lagrangian tracer transport schemes do not conserve heat, water, and chemical tracers, thus making them unsuitable for seasonal to climate purposes or for any tracer studies.

Lines 77-79 – It is precisely for the same reasons that many scientists advocate that one should close the energy budget of ocean models, which cannot be done without retaining the non-conservation of heat in the temperature equation of ocean models, e.g., Tailleux (2010), Dewar, Shoonover, McDougall and Klein, *Fluids*, 2016, https://doi.org/10.3390/fluids1020009. The authors make the implicit assumption that models that treat their temperature variable as conservative but not their total energy are more reliable than the models doing the opposite, i.e., treat their total energy as conservative but not their temperature variable. Shouldn't this be left as an open issue for the community to think about?

Ocean models write their temperature evolution equation so that it is conserved in the ocean interior. Also, as our manuscript points out, and as we showed in pages 20-28 of our response to this reviewer's first review, https://doi.org/10.5194/gmd-2020-426-AC1 , Total Energy is not a conservative variable (and so should not be treated as such when constructing an ocean model in the future).

In more detail, ocean models have, to date, assigned to both their temperature and salinity variables the following two properties,

1. the "potential property", and
2. the "conservative property".

The "potential property" means that when the pressure acting on a fluid parcel changes, without any exchange of heat or salt (that is, during an adiabatic and isentropic pressure change), then the property remains unchanged. The "conservative property" means that when two fluid parcels are mixed, the total amount of that property in the original two fluid parcels is the same as the amount of that property in the final mixed sate. Total Energy is not a potential variable, since an adiabatic and isentropic change in pressure alters the Total Energy of a fluid parcel. Nor does Total Energy possess the conservative property, because both contraction-on-mixing and wave processes allow the internal energy (and Total Energy) of the final mixed fluid parcel to be different to the sum of the internal energies (and Total Energies) of the two original fluid parcels.

For a variable to possess the "conservative property", it is not sufficient that the material derivative of that property is given by the divergence of a flux. Rather, what is needed is that the material derivative of a conservative variable must be equal to the divergence of a flux that is zero in the absence of mixing at that location. That is, the flux whose divergence appears on the right-hand side of the evolution equation of a conservative variable must be a diffusive flux (whether a molecular or a turbulent type of diffusive flux). This feature allows one to integrate over a region in which a mixing event is occurring and be confident that there is no flux through the bounding area that lies outside of the fluid that is being mixed. This is not possible for Total Energy, because even when integrating out to a quiescent surface that encloses an isolated patch of turbulent mixing, the flux divergence term $-\nabla\cdot\left(P\mathbf{u}\right)$ can still be non-zero there.

Note that the volume integral of Total Energy, integrated over the volume of the global ocean, only changes due to the surface boundary conditions (heat fluxes, and fluxes due to precipitation, evaporation, and changes in atmospheric pressure) at the sea surface and at the ocean floor. Indeed, the fact that the material change of Total Energy is caused by the divergence of various fluxes is simply because of the way that the Total Energy evolution equation is constructed (see Appendix B of IOC et al., 2010). That is, it is a direct result of our anthropogenic assertion that "Total Energy cannot be created or destroyed"; but note that it can be transported. This property does not however bestow on Total Energy the conservative property, simply because one of the fluxes that appear on the right-hand side of the Total Energy equation is not a diffusive flux and does not go to zero in the absence of mixing processes. Not only does contraction-on-mixing contribute to $-\nabla\cdot\left(P\mathbf{u}\right)$, but this term also has contributions from various wave processes.

Line 97 – I don't understand the term `contradictions' being used here. Anybody else woulddescribe the approximations made as resulting in errors/uncertainties in need of being quantified, not contradictions. The terms neglected have been shown to be small, and therefore consistently neglected by ocean modellers. Even if one agreed to use the term 'contradiction' here, logic would dictate that resolving the contradiction would be to use the correct equation for potential temperature. Arguing that one should switch to conservative temperature may be a viable alternative, but it is not the logical choice that follows from saying that neglecting some terms in the potential temperature is illogical or contradictory, since the contradiction is eliminated by retaining the terms that the authors say one should not neglect.

We retain the text as is, since to interpret the temperature variable of an EOS-80 ocean model as being potential temperature means that a different heat flux enters the ocean than leaves the atmosphere. This is contradictory, and it affects a quantity (the air-sea heat flux) that is of immense importance for modelling climate. An error in an equation of state, or the absence of source terms in a salt or temperature budget have the nature of approximations to ocean physics (similar to the choice of an uncertain turbulent diffusivity), but to make an (avoidable) error in the driving air-sea heat flux of an ocean model is a more basic and more embarrassing physical error that, in 2021, we should no longer countenance.

Line 99 – May be the 'contradictions' that the authors refer to have been ignored because they are not real contradictions, and just simply because the terms neglected are so small that retaining them would not make any difference, which would be the natural thing to discuss.

As we show in the manuscript, the damage that is done to the air-sea heat flux at a given horizontal location by the interpretation that the temperature variable of an EOS-80 ocean model is potential temperature is not small in comparison to the globally averaged rate that our planet is being anthropogenically warmed. That is, in regions that are comparable in area to an ocean basin, a heat budget analysis using EOS-80 and potential temperature would find a false trend as large as the globally averaged rate that our planet is warming.

Lines 103-104 – 'at the cost of introducing problems elsewhere'. This seems a very strange line of reasoning to me, as the analysis of the problem clearly reveals that the problem that the authors discuss can easily be corrected without introducing any problems elsewhere, by using the correct equation for potential temperature instead of the approximate one. I find it hard to understand why the authors find it worthwhile to discuss inferior solutions.

As we show in the manuscript, it is not possible to accurately formulate the correct air-sea flux of potential temperature, nor is it possible to accurately include the interior non-conservative source term for the potential temperature evolution equation.

Lines 109-111 – "For example, the insistence that a model's temperature variable is potential temperature involves errors in the air-sea heat flux in some areas that are as large as the mean rate of current global warming'' – This is just not true, because the problem that the authors raise is not due to using potential temperature as such, but with not using the exact evolution equation and boundary conditions for it. As shown by Tailleux (2015), correcting the equation for potential temperature to address the authors' criticism would be very easy to implement. It is misleading for the authors to put the blame on potential temperature, where the blame lies in fact with ocean modellers not using the most accurate equation for potential temperature. Potential temperature is a great temperature variable, which has the properties that it has, and there is a priori no problem in using it if the correct evolution equation and boundary conditions are used. IOC et al. (2010) clearly misunderstands this. The best practice is not using Θ instead of θ. The best practice is to use the most accurate evolution equations and boundary conditions, regardless of which variable is used, both variables being perfectly acceptable if used consistently.

Again, it is not possible to accurately do as the reviewer suggests here, and this is why IOC et al. (2010) adopted Conservative Temperature as the temperature to be pursued in ocean models. We show in the manuscript, it is not possible to accurately formulate the correct air-sea flux of potential temperature, nor is it possible to accurately include the interior non-conservative source term for the potential temperature evolution equation. It is hard to see any advantage in attempting to bandage up potential temperature when the adoption of Conservative Temperature is the much cleaner approach. The evolution equations for potential temperature, for Conservative Temperature and for entropy were derived and compared in McDougall (2003) and in Graham and McDougall (2013), and there are clear benefits in adopting Conservative Temperature in ocean models. The choice to adopt Conservative Temperature was made well after the evolution equations for potential temperature and Conservative Temperature were known, and it was made after considering both options (see the discussion of this point in McDougall, 2003).

Line 165 – 'has a mean non-conservation error'. Why do the authors call it an 'error'? The non-conservation of any temperature variable is a real physical process as far as I am aware.

The "error" refers to the error made when the non-conservative terms are ignored.

Line 165 – The authors need to say that the number of '0.3 mW m-2' relies on neglecting the non-conservation of potential enthalpy arising from the Joule heating due to the viscous dissipation rate, and that if the latter was retained, this number would be much larger and not that different from that for potential temperature.

Thanks. We agree and this has now been added.

Line 178 – As pointed out in my first review, the viscous dissipation must balance the mechanical power input by winds and tides, which provides a useful sanity check. 3 TW is a very lower bound for this, which amounts to 10 mW/m2, which is more than 3 times larger than that of Graham and McDougall, 2013. The Graham and McDougall 2013 estimate is therefore implausibly too small.

Thanks. We agree that the Graham and McDougall (2013) estimate of the dissipation of kinetic energy is too small, as it doesn't adequately include the dissipation in the upper ocean. The text has been changed accordingly.

Line 181 – Potential enthalpy had been in used as the thermodynamic variable used in the GISS model, as pointed out in my first review. It would seem justified to cite Russell et al. here, and point out that the variable has been in use way before McDougall (2003) re-discovered it.

The use of potential enthalpy in ocean models by Russell et al (1995) was mentioned as far back as the original theoretical paper of McDougall (2003).

Line 195 – Why is potential temperature not conservative?
This paragraph seems to mix up the a priori unrelated issues of 'conservativeness' and 'how to define heat content'. For clarity, it would be best to discuss the problem of how to define heat somewhere else, since the two issues are only indirectly related. I also find it strange that the section title only mentions potential temperature, given that the physical reasons why 'heat'-like variables are non-conservative are a priori the same for θ and Θ, so why leave Θ out of the section title?

As to the explanation for non-conservativeness, the simplest in my view is to say that both θ and Θ are non-conservative because:

- Neither θ nor Θ mixes linearly (under diffusive effects alone), i.e., the θ or Θ of the mixture of two water samples is different from the mass weighted average of the two samples.
- In a turbulent ocean, θ and Θ also systematically increases during mixing events due to turbulent dissipation of kinetic energy by viscous processes.

Then, the conditions can be separated for θ and Θ. For instance, the authors could say that the non-conservativeness of Θ is controlled by the temperature dependence of T/θ whereas the non-conservativeness of θ is controlled by the temperature dependence of $T\ c_{pr}/\theta$. – If c_{pr} were assumed constant and T/θ were a function of pressure only, as for a dry atmosphere, then both CT and θ would be considerably more conservative than in seawater and would have identical degree of non-conservativeness. The fact that θ is less conservative than Θ is due to the temperature dependence of c_{pr} – with a lesser role forthe salinity dependence. This is shown by Eqs (23) and (25) of Tailleux (2010), which I think the authors should refer to. The method developed by Tailleux (2010) (or Tailleux (2015)) is the most general currently available and is valid for the full Navier-Stokes equations. At the moment, this method is the one that underlies the construction of energetically consistent approximations. The methods discussed in IOC et al. (2010) and Graham and McDougall (2013) are much less general. Moreover, they fail to incorporate viscous dissipation as part of the definition of non-conservation of θ and Θ.

This discussion is incorrect on multiple grounds. First, Total Energy is not a conservative variable (see above) and so the derivation of Tailleux (2010, 2015) is flawed; that is, it is energetically inconsistent, as shown in our response to the reviewer's first review of this present paper. Second, it is not correct to say that McDougall (2003) and Graham and McDougall (2013) ignore the contribution of the dissipation of kinetic energy to the production of potential temperature and Conservative Temperature. This term appears in the relevant evolution equations in those papers.

Line 215 – "This suggestion has been made, for example, by Tailleux (2015)''.
First, I don't think that the suggestion has been made by anybody else. Second, the method proposed by Tailleux (2015) is merely to make use of the passage relations $\theta = \theta(S_A, \Theta)$ and $\Theta = \Theta(S_A, \theta)$ to reformulate the evolution equation for Θ used by a TEOS-10 model to obtain a mathematically equivalent one but for potential temperature. In other words, the evolution equation for potential temperature can be obtained by a simple change of variables from that for conservative temperature. Alternatively, one could also diagnose the non-conservative terms in the evolution equation for potential temperature to close the energy budget of the EOS-80 numerical ocean model considered, as in Tailleux (2010). Both strategies circumvent the difficulties raised by the authors and show that improving the equation for potential temperature would be a trivial exercise.

Given that both Tailleux (2010, 2015) have proposed concrete solutions to compute the non-conservative terms to be added to the evolution equation for potential temperature, I find it odd and rather non-collegial for the authors to assert that such approaches would be unworkable. If the authors do not understand how to improve the evolution equation for potential temperature, this does not mean that is necessarily true of everybody else. Instead of unfairly disparaging Tailleux's work, the authors could simply say that Tailleux's suggestions remain to be implemented and tested and compared with a Θ-based formulation.

We stand by what we have written, and if further detail were needed it can be found in our response to the reviewer's first review, at https://doi.org/10.5194/gmd-2020-426-AC1. Again, it is not possible to accurately do as the reviewer suggests here, and this is why IOC et al. (2010) adopted Conservative Temperature as the temperature to be pursued in ocean models. We can see no advantage in attempting to approximately bandage up potential temperature when adopting Conservative Temperature is the much cleaner approach.

Lines 239-242 – I think that the authors misunderstand and misrepresent Tailleux (2010, 2015)'s approach. Indeed, Tailleux (2010)'s approach is fully deductive and rigorous, contrary to what the authors seem to suggest. Specifically, Tailleux's approach to obtain a mathematically explicit expression for the non-conservation of Θ an θ is identical to that used by Prigogine and the Belgian school of non-equilibrium thermodynamics (improved by Lesley Woods, 1975) to obtain a mathematical expression for the non-conservative production of entropy. Physically, this approach consists in defining the non-conservation of specific entropy (and by extension that of Θ or θ) as what is needed to make total energy conservative as per the law of energy conservation. Here, the term 'conservative' means that all the terms entering the evolution equation for total energy can be written as the divergence of a flux, which is the usual definition. In their paper, the authors seem to confuse the term 'conservative' with the property of 'mixing linearly', as when they say total energy is not conservative, they clearly mean that total energy does not mix linearly. Saying that total energy is non-conservative is very confusing.

Now, the full evolution equation for the specific enthalpy in seawater takes the form:

$$\frac{Dh}{Dt} = -\frac{1}{\rho}\nabla \cdot \left(\rho F_q\right) + \upsilon\frac{Dp}{Dt} + \varepsilon_k$$

(see Eq. B19 of the latest version of TEOS10 manual with remineralization term removed). In a turbulent ocean, neither the pressure term nor turbulent viscous dissipation can be neglected, so Graham and McDougall's (2003) assertion that the locally referenced potential enthalpy mixes linearly if one neglects viscous dissipation is inconsistent with the fact that pressure always fluctuates in a turbulent ocean. However, it appears to be true that the first-principles expressions for the non-conservation of Θ and θ obtained by Tailleux (2010) can also be obtained by treating specific enthalpy as if it were linearly mixed, i.e., by omitting the pressure term in the above equation. However, because Tailleux (2010) and Tailleux (2015) are rooted in a fully deductive and first-principles approach, which is not the case of Graham and McDougall (2013), the correct way to justify the assumption made by Graham and McDougall (2013) is by showing that it follows from the exact results of Tailleux(2010), not the reverse.

We do understand the approach of Tailleux (2010, 2015) and we have shown that these papers have made errors in the development of their equations. In the Appendix of our response to the reviewer's first review at https://doi.org/10.5194/gmd-2020-426-AC1 we show that Total Energy is not a conservative variable (as we have defined this term above). The reason why Tailleux (2010, 2015) thought that Total Energy is a conservative variable is that these papers ignored the non-conservative nature of the $-\nabla \cdot \left(P\mathbf{u}\right)$ term. Rather, as in well known in the thermodynamic literature, when fluid parcels mix together at a certain pressure, the thermodynamic quantity that is conserved is enthalpy. This is the basis of the McDougall (2003) and Graham and McDougall (2013) papers.

Lines 253-256 – 'However, these expressions are written in terms of molecular fluxes, and it is not possible to use these expressions to evaluate the non-conservation in a turbulently mixed ocean'. I really don't understand where does this come from. Again, the authors appear to assume that because they do not know how to do something, this should also be the case of everybody else, which seems to me to go against the collegial nature of science. Moreover, the Navier-Stokes equations are well accepted to describe both laminar and turbulent motions, so clearly Tailleux (2010,2015)'s expressions pertain to a turbulently mixed ocean, contrary to what the authors say. What is true, however, is that the expressions remain to be linked to turbulent fluxes or microstructure measurements in order to allow for their evaluation. One way this could be done is by using expressions such as the Osborn-Cox model linking the dissipation of temperature variance to the turbulent heat diffusivity as follows:

$$\kappa \frac{|\nabla T'|^2}{\left(\frac{d\overline{T}}{dz}\right)^2} = K_T$$

On the left-hand side, the terms involve the molecular fluxes of temperature as well as the mean temperature profile, whereas on the right-hand side appears the turbulent diffusivity for the mean temperature. The authors' remarks have incited me to rework on the issue in order to show that such expressions can indeed be linked to microstructure measurements and evaluated from first principles, as I hope to show in a forthcoming publication.

The referee now seems to agree that the approaches of Tailleux (2010, 2015) "remain to be linked to turbulent fluxes ...". But, the McDougall (2003) and Graham and McDougall (2013) approach *has already been written in terms of the turbulent fluxes*. This is what several sections of those papers do. See for example, Eqn. (38) of Graham and McDougall (2013). That is, this link to turbulent mixing that referee discusses in this comment has already been done, and has been published in 2003 and in more detail in 2013.

Line 331 – What this describes is 'density salinity' – My understanding is that density salinity is always different from absolute salinity except when all the haline contraction coefficients for each of the chemical constituents are identical. May be this can be mentioned and commented upon.

The Absolute Salinity of TEOS-10 is defined to be its "density salinity"; see section A.4 of IOC et al. (2010), the paper by Wright et al. (2011) and page 169 of the summary paper of Pawlowicz, McDougall, Feistel and Tailleux (2012). This definition of Absolute Salinity was adopted primarily for two reasons. First, specific volume is the thermophysical quantity whose sensitivity to the variations in seawater composition has the most impact on ocean and climate circulation and fluxes. Second, it is possible to measure the specific volume of a liquid in an SI-traceable manner, so providing a link between Absolute Salinity and SI-traceability (such an SI-traceable route to the measurement of Practical Salinity had been lacking prior to the adoption of Absolute Salinity by TEOS-10).

Line 341 – Can the authors clarify whether the relation is actually between S* and SA, or between S* and SD (density salinity).

The Absolute Salinity is the same as "density salinity"; see the previous response above.

Lines 361-367 – Can the authors comment on the differences in computational efficiency of the equation of state between the Jackett and McDougall (1995) and Roquet et al. (2015). This information is important for ocean modellers to decide whether to switch or not to switch.

The equation of state of Jackett and McDougall (1995) is a rational function whereas that of Roquet et al. (2015) is a straightforward polynomial. While they both are approximately equally as computationally efficient for the evaluation of specific volume, the Roquet et al. (2015) is substantially more efficient when the partial derivatives (with respect to SA and CT) are calculated. Computer codes actually do more evaluations of these partial derivatives than they do of specific volume itself (for example, in the neutral physics part of the code), so the Roquet et al. (2015) form of the equation of state comes with this computationally advantage (as well as being a function of the modern salinity and temperature variables).

Lines 381-392. I agree that TEOS-10 has conclusively shown that variations in composition potentially matters for estimations of the thermal wind. However, it is also essential that the equations of motion used by ocean models be mathematically and physically well posed. As far as I understand the problem, while it is obvious that the equations of motions based on the use of reference composition salinity are well posed, it seems to me that this is not the case if we use absolute salinity (or rather density salinity). Moreover, as well as making the equations of motion ill posed, the use of density salinity also seems to screw up the energetics by introducing spurious sinks and sources of energy. As far as I am aware, TEOS-10 never wrote down a mathematically consistent set of equations based on the use of absolute salinity. I would very much like to see the qualitative considerations about the importance of the variations in composition accompanied by the authors writing down a full set of equations of motion that can be studied by mathematicians and dynamicists like me. If the authors cannot produce a mathematically well posed set of equations using absolute salinity, it seems to me that they should not promote it as a meaningful basis for ocean modelling. If model equations using absolute salinity are ill posed as I think they are, the consequence is that it is a priori impossible to be sure of how to interpret the results of McCarthy et al. (2015). To me, this is a key issue that the IOC et al. (2010) and the authors appear to have overlooked.

No ocean model has yet been published that has included the non-conservative source terms in its evolution equation for salinity. In section A.20 of IOC et al. (2010) we described how we thought it should be done. But only when an ocean modelling group decides to adopt this suggestion (or modifies it) will we, as a global community of oceanographers, learn more about this. The reviewer talks of the non-conservation of energy (but, which type of energy?). Perhaps more basic than this is the non-conservation of Absolute Salinity. Then, at some stage we will need to be adding the non-conservative dissipation of kinetic energy to the right-hand side of the CT evolution equation in ocean models. And there is yet another way that energy is not conserved in an ocean model. This is due to the fact that in the ocean interior, the most consistent interpretation of the SA and CT in an ocean model is as the thickness-weighted mean values of SA and CT (see the Temporal-Residual Mean papers). The difference between these TRM values of SA and CT and the Eulerian-mean versions also leads to the non-conservation of energy, as pointed out by Peter Killworth. Indeed, the non-linear nature of the equation of state of seawater causes many conceptual issues that we are still learning how to deal with; such as thermobaricity, cabbeling, the ill-defined nature of neutral surfaces etc.

Ocean modelling groups are not keen to add non-conservative source terms to either their SA or CT evolution equations, since this would deny these groups, and the many other folk who analyze model output, the ability to evaluate things like the meridional freshwater flux or the meridional heat flux; how would one include the effects of the non-conservative source terms in these calculations?

Line 406 – What is the way to compute Cp(S*,theta,0) using the TEOS10 software? Can they provide the appropriate lines of code that would need to be invoked to compute it?

Since the models do not carry any non-conservative source terms in their salinity evolution equation, and any sea-surface salinity restoring boundary condition restores to S* (which is virtually the same as SR and SA at the sea surface), the GSW software should be called with the model's salinity variable (if it is a TEOS-10 ocean model) and with u_PS times the model's salinity if the ocean model is a EOS-80 based model. So the GSW software call should use the code **gsw_cp_t_exact** with the arguments being S*, potential temperature and zero sea pressure.

Line 412 – The fact that the authors use sometimes S*, sometimes SA is confusing.

We trust this is clear now.

Lines 414-415 – May be add a physical explanation for why the temperature or rain is not treated consistently.

The interested reader can read about the difficulties that atmospheric models have in this regard by consulting the references given.

Lines 444-445 – That's the authors interpretation. The alternative and more common interpretation is that these errors are accounted for in the estimation of errors and uncertainties affecting the simulated θ field.

We have made the point that having the ocean receive a different heat flux than the atmosphere gives it is simply not acceptable in 2021. It is embarrassing that as a community we have lived with this discrepancy for the past century. Fortunately, since the adoption of Conservative Temperature and TEOS-10, there is no longer any need to put up with this incorrect physics in our coupled models.

Lines 453 – To ensure that the model equations are well posed, many ocean models will assume that their salinity argument is reference salinity rather absolute salinity. Again, I have yet to see a consistent set of equations based on absolute salinity.

The reviewer seems to have the properties of these salinity variables back-to-front. If an ocean model were to carry Reference Salinity as its prognostic salinity variable, it would need to also carry the non-conservative source terms that represent to biological effects on Reference Salinity. It is Preformed Salinity that obeys a conservation equation (with no source terms), whereas Reference Salinity and Practical Salinity at depth in the ocean are the result of not only advection and turbulent mixing processes, but also of the non-conservation of these variables due to the dissolution of sinking fecal pellets. This is explained in IOC et al. (2010), Wright et al. (2011) and Pawlowicz, McDougall, Feistel and Tailleux (2012).

Lines 481-488 – I think that this paragraph is going to cause considerable confusion in the community as it seems inconsistent with the way things have been described before. First, IOC et al. (2010) says that the new equation of state is defined in terms of absolute salinity (while in fact using density salinity to estimate absolute salinity, even though the two are supposed to be somewhat different). Now, the authors appear to say that it is defined in terms of preformed salinity S*, which is always numerically different from absolute salinity. Does that mean that the authors are actually already moving away from the recommendations of TEOS-10? Nothing of what the authors say about salinity in this paper makes any sense to me. I just don't understand where all this come from, and I suspect I won't be the only one. I think that it would greatly help if the authors could write the model equations that ocean modellers are supposed to solve with the proposed interpretation, may be in an appendix.

We are sorry that the reviewer finds these salinity issues confusing, and yes, we agree that he is not alone in this. We trust that the explanations given above help. The key papers regarding the several salinity variables are the papers by Pawlowicz (2010), Pawlowicz, Wright and Millero (2011), section A.4 of IOC et al. (2010), the paper by Wright et al. (2011) and the summary paper of Pawlowicz, McDougall, Feistel and Tailleux (2012).

Lines 489-492 – What is this based on exactly?

It is based on the fact that (1) the salinity variable in an ocean model is treated as a conservative variable, and (2) that this variable is restored at the sea surface in the same way that is Preformed Salinity (which is equal to both Reference Salinity and Absolute Salinity there). These are the features of Preformed Salinity. Hence the salinity variable that these models have been carrying all these years is Preformed Salinity, even if previous ocean modellers have thought that their models carried a different type of salinity. Just because there are hundreds or perhaps thousands of published papers saying that the model output is Practical Salinity does not mean that their interpretation is correct.

Lines 499-503 – "The model's salinity variable will drift towards being preformed salinity." I really don't understand why. How can the authors make such an assertion without substantiating it? For instance, could the authors write down an evolution equation for the drift that would clarify the relaxation time scale and convince us that what the authors describe has a counterpart in the mathematical world?

This assertion of ours is based on the fact that (1) the salinity variable in an ocean model is treated as a conservative variable, and (2) that this variable is restored at the sea surface in the same way that is Preformed Salinity (which is equal to both Reference Salinity and Absolute Salinity there). These are the features of Preformed Salinity. Hence the salinity variable that these models have been carrying all these years is Preformed Salinity, even if previous ocean modellers have thought that their models carried a different type of salinity.

Lines 536 – This is not how models work. Indeed, as far as I understand the issue, the temperature variable used by a model is not a matter of interpretation, it is a matter of declaration. The first step in constructing a model is to declare what its dependent variables should be. Once the variables have been declared, the next step is to decide on the evolution equations and boundary conditions that one will use to describe their temporal evolutions. To me, it is essential that models be based on precise definitions and declarations, not interpretations, so that what we do can be easily understood by our colleagues mathematicians and atmospheric scientists. I am pretty sure that mathematicians cannot understand what the authors mean by 'interpretation', which is bound to leave them very confused. My impression is that the authors use the term 'interpretation' because they want the reader to accept their view that the physical meaning of the variables used by an ocean model is open to discussion, which seems questionable at best.

We totally disagree with this comment (and previous such comments). How could the temperature or salinity variable of an ocean model be what a person simply declares it to be? One could for example declare that the model's temperature variable is ten times the in situ temperature in degree F. Such a declaration does not make it so. Rather, the model's temperature and salinity variables must be interpreted in terms of how these variables are forced and transported in the model.

We have realized that in the case of temperature, we must not allow the ocean to receive more (or less) heat than the atmosphere delivers, and this means that it is no longer acceptable to interpret the temperature variable in EOS-80 models as being potential temperature. That this conflicts with the hundreds or perhaps thousands of published papers on ocean models does not invalidate our revised interpretation of the temperature variable of these models. This responsibility of a scientist to interpret what the model's variable is, should not be controversial.

Line 543 – I disagree that this is a conclusion. It looks much more like an opinion or assertion. It would be useful if the authors could provide the reader with some experiments to run that would enable the ocean modelling community to test its validity.

These lines of the manuscript summarize our most prominent result, and how it differs from what one reads in Griffies et al. (2016). We are disappointed that the reviewer does not agree with the thesis of our paper.

Lines 556-557 – Again, my view is that the evolution equation and boundary conditions have no bearing on the definition of a physical quantity. For this statement to be acceptable, one has to accept that the definition of a physical quantity is not independent of its assumed evolution equation and boundary conditions, in contrast to what is generally done (as far as I understand it).

We totally disagree with this comment (and the previous such comments). See our discussion above, for example, in response to the reviewer's comments on line 536. Contrary to what the reviewer says, it is indeed crystal clear that the interpretation *"of a physical quantity is not independent of its assumed evolution equation and boundary conditions …".* How could it be otherwise? For example, if an ocean modeler decided to "define" their model's temperature variable to be the number of moles of human growth hormone per kilogram of seawater, this definition does not make it so. Rather, the temperature variable that comes out of the model is a result of how the variable called temperature is treated inside the model (initial conditions, boundary conditions and evolution equation).