Response to Reviewer Remi Tailleux (RT)

Reviewer RT has provided us with a lengthy and strongly worded disagreement with our manuscript. There are many aspects to this disagreement (he provides some general comments, followed by 4 main points of contention, and 22 specific comments), which we will address below. RT's comments are inserted in black text, and our responses are in blue text. A number of the specific comments, ("specific comments" numbered 4, 9, 10, 11, 17, 20 and 22) encourage us to re-work our analysis by taking Total Energy, $\mathcal{E} = u + \frac{1}{2}\mathbf{u} \cdot \mathbf{u} + \Phi$ to be a conservative variable. These comments of the present review by RT, as well as the Tailleux (2010) and Tailleux (2015) papers are based on the assumption that Total Energy is a conservative variable, but this assumption is incorrect. Because this issue underlies so many of the reviewer's comments, and is also the basis of the Tailleux (2010) and Tailleux (2015) papers, we address it thoroughly in the Appendix of this Response where we prove that Total Energy is not a conservative variable.

We initiated the present research by asking whether it was possible to ensure that coupled climate models do not lose heat at the air-sea interface, since the usual assumption that an EOS-80 based model carries potential temperature as its temperature variable means that not all of the heat that leaves the atmosphere arrives in the ocean. Fortunately, as discussed at length in our manuscript, this issue is solved by simply interpreting the prognostic temperature variable in EOS-80 based ocean models as being Conservative Temperature. We show that doing so means that the equation of state in the EOS-80 based ocean model is not as accurate as it could be, but there are also many other aspects of ocean models that we know are not perfect; just think about the selection of diffusion coefficients, and the temporal drift of deep ocean temperatures. These are all aspects of our science that we oceanographers continue to work on, but at least let us not continue to lose heat at the air-sea interface, especially when in our paper we describe a very easy fix.

The "general comments" appear to consist mostly of opinions, and so we do not reproduce them here in full. However, four statements by RT in particular are key, since we have different views on each which we feel it is important to highlight:

> (1) From a theoretical viewpoint, the identity and nature of the temperature and salinity variables used by a numerical model must be consistent with the choice of equation of state.

Although this statement is reasonable on the face of it, the fact is that our existing equations of state are not exact models of actual seawater, as RT does recognize. Thus, it is not a question of "consistent" versus "inconsistent" – instead all choices have some degree of inconsistency. Given that fact, this paper is then motivated by our insight that the tradeoff incurred by adding a little more inconsistency in the equation of state may be worthwhile, if some other advantage can be gained for a particular purpose – and, in particular, for numerical modelling and CMIP analysis, if we can take advantage of some of the strengths and features of current GFD numerical computational schemes without modifying them (incidentally, we feel that this approach is in fact directly in line with RT's later comments about the superiority of considering "approximations" rather than the true/false dichotomy he apparently attributes to us. We are looking for "better", not necessarily "ideal"). So while

our thesis will involve a little extra error in the equation of state of EOS-80 based ocean models as we have discussed in the article, at least we have provided an easy fix to one of the embarrassing aspects of these EOS-80 based climate models, namely that some of the air-sea heat flux disappears.

An ocean model contains many moving parts, including the surface and sea-floor boundary conditions, the equation of state, and the ways in which interior mixing processes (of which there are many) are parameterized. So, it is clear that the equation of state is but one aspect of an ocean model. It so happens that the air-sea flux condition employed in EOS-80 based ocean models is thermodynamically inconsistent with treating the model's temperature variable as being potential temperature, but is thermodynamically consistent with interpreting the model's temperature variable as being Conservative Temperature. Hence our paper has explored the consequences of interpreting an EOS-80 based ocean model's temperature variable as being Conservative Temperature. We find that this is a viable choice, and very importantly, it means that the heat fluxes that are exchanged between the atmosphere and ocean occur without the loss of heat. We are sure that most scientists would agree that it is not advisable to lose heat at the air-sea interface when modeling climate. The point is that the equation of state is but one aspect of an ocean model, and of a coupled climate model, in which the model's temperature variable appears. The combination of all the moving part of the coupled model work together to determine what the ocean model's temperature actually is; not just the equation of state.

In addition, the general comments also contain the inaccurate contention that:

> (2) "Encouraging ocean modellers to test the impact of using Absolute Salinity without telling them about its potential detrimental effects seems a bit misleading".

A key point about TEOS-10 is that its makers realized that no single "salinity" variable can meet all possible needs, and so rather than attempting to define a "jack of all trades, master of none" type of salinity, several different ones were devised – Preformed, Reference, Absolute, Solution, etc., to provide maximum effectiveness for different purposes. The TEOS-10 Absolute Salinity is in fact designed to provide estimates of specific volume at highest precision, after taking into account spatial changes in the relative composition of sea salt, and is therefore primarily aimed at observationalists, rather than at numerical modellers. Reference Salinity is designed to match best with conductance-based measurements, and hence is also not aimed at numerical modellers. Both of these two forms of salinity are subject to internal sources and sinks from ocean biogeochemical processes.

It was the viewpoint of the SCOR/IAPSO WG-127 that oceanographers were (as always) free to use lower precision salinity variables if that suited their needs, but the consequences of these choices could now be better understood by comparison with the "best" possible variables for a particular purpose (these points and their rationale are explained at length in Wright et al. (2011), and, incidentally, many of the shortcomings of TEOS-10 have been considered, described, and enumerated in the numerous publications that underly this standard). Conservative Temperature is also not touted as an "ideal" parameter, merely a "much better" parameter than potential temperature for certain purposes, e.g., maintaining a conservative-under-mixing behavior in the current ocean.

Note also that the samples whose measured specific volumes were incorporated into both the EOS-80 and TEOS-10 equations of state were of Standard Seawater whose composition is close to Reference Composition. Hence, the EOS-80 and TEOS-10 equations of state were actually constructed with Preformed Salinity as their salinity arguments, not Reference Salinity. That is, for a seawater sample that is not of reference composition, calling the TEOS-10 expression for specific volume with Reference Salinity as the salinity argument will not give an accurate expression for specific volume; the salinity argument should be Absolute Salinity. And for an ocean model that has no non-conservative interior source terms in its salinity evolution equation, and is initialized at the sea surface with Preformed Salinity, then the only interpretation for the model's salinity variable is Preformed Salinity, and the use of the TEOS-10 equation of state will then yield the correct specific volume.

All ocean models treat their salinity as being a conservative variable, and they also all initialize this salinity at the sea surface as Reference Salinity (or, equivalently, as Practical Salinity). At the sea surface the concentration of nutrients is small and so Reference Salinity at the sea surface is virtually the same as Preformed Salinity (and to Absolute Salinity). Since these models are initialized (and restored) to surface values of Preformed Salinity, and since both the models and Preformed Salinity are conservative, then the output salinity of these models has only one interpretation, named Preformed Salinity. The manuscript makes a clear case for this, and this review by RT has not mounted a case against this interpretation. Yes, this is different to what Griffies et al. (2016) recommended. We are not able to change what is published in that paper, but we can push forward with the science and accept the compelling arguments that arise.

The reviewer also raises an issue concerning the general utility of Conservative Temperature vs. potential temperature, wishing that a different paper, a modelling paper that

> (3) replace[d] the boundary condition Q/c_p by replacing the currently constant heat capacity by $c_p(\theta, S, p_a)$, and by adding the missing non-conservation terms by diagnosing these in the manner proposed by Graham and McDougall (2013) or Tailleux (2015).

had been written by us instead of the present one under review. Now, while the utility of this task that the reviewer describes can be debated, it is our view that undertaking this task is not as straightforward or as interesting as the reviewer suggests, for reasons that we explain in the following paragraphs.

The recommendation of IOC, SCOR and IAPSO that ocean models switch from using potential temperature to using TEOS-10's Conservative Temperature was made after careful consideration of many factors. However, the reviewer is pushing the thesis of Tailleux (2015) (see especially section 5 of that paper) that ocean models can be formulated just as well in terms of potential temperature as they can be in terms of Conservative Temperature. We disagree, and here we summarize why the path suggested by Tailleux (2015) and by the present review of our manuscript, is impractical and unworkable.

The adoption of Θ overcomes the following four rather serious disadvantages of adopting the approach advocated by Tailleux (2015) and by the present review, namely

1. It is not possible to accurately choose the value of the isobaric heat capacity at the sea surface that is needed when θ is the model's temperature variable. The problem arises because of unresolved spatial and temporal variations in the sea surface salinity (SSS) and SST [for example, unresolved rain events that temporarily lower the SSS but are not represented in the time-averaged data]. These unresolved variations in SSS and SST act in conjunction with the nonlinear dependence of the isobaric specific heat on salinity and temperature to mean that it is not possible to obtain the appropriately averaged value of the isobaric specific heat.

2. It is not possible to accurately estimate the non-conservative source terms for θ. These terms are the product of a turbulent flux and a mean gradient, and in an eddy-resolved ocean model, how would one go about finding the eddy flux of θ, which depends on how the averaging is done in space and time. [How to calculate the appropriate mean gradients, over what space and time scales, and how to treat non-divergent eddy fluxes?]

3. Calculating the meridional heat flux through an ocean section cannot be done accurately if θ is the model's temperature variable. Because of the interior source terms, in order to calculate the heat flux through an ocean section, one would presumably need to abandon θ and do a conversion to Θ, and then evaluate its transport across the section. In this way, one would have gone full circle, back to treating Θ as the "heat-like" variable whose transport can be compared to the air-sea heat fluxes. So why not adopt Θ as the model's variable to start with? But there is a more basic point here as well. Normally the meridional or zonal heat flux across sections is done with the monthly or annual mean properties, and the conversion from one temperature variable to another cannot be done accurately when the salinity and temperature vary in space and time due to the nonlinear dependence of Θ on salinity and θ (we discuss this effect in Figure 9 below). These issues are avoided when Θ is the model variable.

4. Ocean modellers often use the conservation of salinity and the model's temperature variable to check the model's numerics. If θ is adopted as the model variable, this is no longer possible because θ is not a conservative variable.

In summary, the reviewer discusses, as does Tailleux (2015), that ocean models could well retain potential temperature θ as the model's temperature variable, rather than adopt the TEOS-10 recommendation of using Conservative Temperature Θ. The above 4 points show that doing so means that (1) the air-sea heat flux cannot be accurately incorporated into the ocean model, (2), the non-conservative source terms that appear in the θ evolution equation cannot be estimated accurately, (3) neither can the ocean section-integrated heat fluxes be accurately calculated, and it seems even the inaccurate method that would be employed to do this involves adopting the TEOS-10 approach of calculating the section-wide flux of Θ, and (4), an important and convenient conservation check that is routinely employed by ocean modelers would not be available.

Therefore, we see no advantage to adopting the approach suggested by Tailleux (2015) (as repeated in this review); rather there are the above four disadvantages. Hence, as SCOR/IAPSO Working Group 127, IAPSO and SCOR recommended, and as adopted by IOC IAPWS and IUGG, by far the cleanest way to do ocean modelling is to adopt TEOS-10's Conservative Temperature Θ as the ocean's temperature variable. Nothing in Tailleux (2015), nor in the arguments of the present review, gives pause to this recommendation of TEOS-10.

Finally, RT questions how we know that Conservative Temperature (CT) is much more conservative than potential temperature,

(4) As far as I can judge, running an EOS-80 based model with the current and corrected equation is the only way to ascertain that the current way of treating potential temperature as conservative is as based as Prof. McDougall claims. It is therefore imperative if we are to settle this issue once and for all that somebody, preferably Prof. McDougall and his team, perform such an experiment, which is the only way I can think of to establish a sound and rigorous physical basis for switching to CT. Failing this, whether or not treating θ as conservative is as bad as Prof. McDougall claims will remain speculative and purely based on indirect evidence. In any case, I don't think that the authors' ideas and recommendations should be published until their scientific merits has been established by running an EOS-80 based model with the "correct" potential temperature equation.

The use of Θ as the model's temperature variable is expected to reduce the diffusive effects of the non-conservative nature of the model's temperature variable by two orders of magnitude, compared with using potential temperature. Of this we can be sure, since there are now at least four studies [McDougall (2003), Graham and McDougall (2013) and Tailleux (2010, 2015)] that show that the non-conservative diffusive source terms for Θ and θ are in the ratio of 1:100 or so. Also, the influence of the dissipation of kinetic energy ε can be added as a source term to the model's Θ equation if and when the model's knowledge of ε at run time is considered reliable, while this addition of ε cannot sensibly be done with the model's temperature variable is θ.

But how large are the errors caused by using θ as the model variable? In the literature we can read at least two ways of answering this question. First, as discussed in the TEOS-10 Manual, IOC et al. (2010), if an ocean model is forced by the air-sea heat flux boundary condition (as opposed to a restoring condition on temperature), the differences between an ocean model run with Θ and θ is simply the differences between these variables, that is, $\theta - \text{gsw\_CT\_from\_pt}(S_A, \theta, p)$, as illustrated in A.13.1, where we see a range of temperature differences exceeding $0.2°C$.

The second way of estimating these differences is as done by Graham and McDougall (2013) where they formed vertical integrals of the non-conservative diffusive source terms, finding that those of Conservative Temperature were a factor of 120 less than those of potential temperature. Expressed in terms of an equivalent error in the air-sea heat flux, Graham and McDougall (2013) found that twice the r.m.s. value of the air-sea flux error when ignoring the non-conservative terms of potential temperature was $\pm 120\,\text{mWm}^{-2}$. This is not a small error, and is best avoided. The area-mean value of this air-sea flux error is smaller, at around $-10\,\text{mWm}^{-2}$, but we oceanographers and climate scientists are concerned not only with the volume integrated heat content, but we also care about the accuracy of regional climate and regional climate projections. Hence $\pm 120\,\text{mWm}^{-2}$ is the relevant error measure, not $-10\,\text{mWm}^{-2}$.

So we do already know the magnitude of the damage done to ocean models by using potential temperature as the model's prognostic variable. It is a factor of 120 larger than the corresponding non-conservative diffusive error that remains when adopting Conservative Temperature as the ocean model's prognostic variable (see Table 1). Further studies as suggested by the reviewer here could indeed be attempted, but as discussed on pages 3-4 above, there are four reasons why the approach advocated by the reviewer is impractical and unworkable, particularly in an eddy-rich ocean simulation.

We now move on to address RT's four "major points of contention and disagreement" and the 22 "specific comments". In each case we fully include the reviewer's comment in black text and we reply to each comment in blue text.

Major Points of Contention and Disagreement

1. **"Interpreting θ as CT" is equivalent to "interpreting an orange as an apple"** The authors' recommendation that θ in EOS-80 models should be interpreted as CT presupposes that the two quantities are of the same nature, but we all know that this is not the case. Indeed, while θ is truly a temperature that can be experimentally measured, CT is truly a non-measurable re-scaled energy quantity disguising as temperature. Moreover, since enthalpy and potential enthalpy are defined up to a linear function of salinity, it follows that the construction of CT involves the specification of three arbitrary parameters, two associated with the said linear function of salinity, one associated with the least-square determination of c_p^0. Now, while Prof. McDougall knows about the values of the three arbitrary parameters determining his construction of CT, the potential temperature of an EOS-80 model obviously does not. Can the authors explain how it is somehow possible for θ to morph into CT without θ having any knowledge of the particular determination of CT it is supposed to morph into? How is it possible for θ to somehow morph into CT if it does not know which determination of CT it is supposed to morph into?

Both the enthalpy and the entropy of a binary fluid (such as we usually suppose seawater to be) are indeed unknown and unknowable up to linear functions of Absolute Salinity. This means that there are four unknown and unknowable constants in the Gibbs function of seawater. While having two unknown constants in the definition of enthalpy sounds as though it might have some undesirable consequences, the fact that these two constants are not only *unknown* but are also *unknowable* means that their values have no consequences. Why? Because, if the values of these two constants had any real-world consequences, then we could measure those consequences and hence determine the values of the constants. But since the constants are unknowable, there can never be any observable consequences. The unimportance of these four arbitrary constants in the seawater Gibbs function is discussed towards the end of Appendix B of the TEOS-10 Manual (IOC et al., 2010).

That is, it is well known from advanced thermodynamic texts that there are not (and cannot be) any consequences of any particular choices that are made for these four constants. TEOS-10 made specific choices for the four constants in the seawater Gibbs function, and the choices were made consistent for the Gibbs functions of ice, freshwater, seawater and humid air.

To repeat, if we were to take the TEOS-10 definition of enthalpy and add to it, for example, $\left(1000 + 10\,S_A\big/\mathrm{g\,kg^{-1}}\right)\mathrm{J\,kg^{-1}}$, we would have a new definition of enthalpy, and no measurement that has been made, or could ever be made in the history of the universe, could ever prefer one definition over the other; they both are correct and indistinguishable from each other in terms of any observable consequence. The same can be said of the two unknown and unknowable constants in the definition of the entropy of seawater; these also have no consequences and cannot ever have any consequences.

Having discussed the four unknown constants in the seawater Gibbs function, we now discuss the fifth arbitrary constant of TEOS-10, namely the value that was chosen for c_p^0. Again, this value is completely arbitrary. SCOR/IAPSO WG127 could have chosen it to

be, for example, $1\,\mathrm{J\,kg^{-1}\,K^{-1}}$ in which case Conservative Temperature would have the same numerical value as potential enthalpy. The equation of state could still be defined in terms of this new version of Conservative Temperature, giving exactly the same values of specific volume. At the sea surface, there is still the need to convert from the model's Conservative Temperature to the SST for the calculation of air-sea fluxes from bulk formula, and this conversion could still be done. Hence, the choice of each of the five constants of TEOS-10 have no impact whatsoever on the forward ocean modelling practices of TEOS-10 as recommended by IOC, SCOR and IAPSO in the TEOS-10 Manual, IOC et al. (2010).

The above comments about the un-importance of the five arbitrary TEOS-10 constants apply to when TEOS-10 is adopted (in its entirety). However, in the present paper we are discussing the messy middle ground where the entire ocean community has not yet adopted TEOS-10, and we are suggesting a way to interpret the output of EOS-80 based ocean models using concepts from TEOS-10. In this situation, three of the five arbitrary constants *are* important.

When it comes to re-interpreting an EOS-80 based ocean model as having Θ as its prognostic temperature variable, three of the five arbitrary constants are important in making the mean values of Θ and θ similar at the sea surface. This is by design. The impact of this design is on the air-sea flux that the model draws down from the atmosphere via the flux bulk formulae, as discussed in the text of our manuscript. The three enthalpy-based constants have each been chosen (two by Rainer Feistel, and one [c_p^0] by Trevor McDougall) to minimize the difference between Θ and θ at the sea surface. If we had not minimized the difference between these temperatures at the sea surface, the air-sea flux arising from the bulk formulae would be more different than they presently are when Θ is used as the SST instead of θ .

As described in the text, this difference in the air-sea heat flux when using the two different interpretations of the model's surface temperature is equivalent to the ocean modeler specifying a slightly different set of bulk formulae; this issue is different to the other ones discussed in our paper which go to the thermodynamic consistency of the heat fluxes in the ocean versus those in the atmosphere.

That is, while it is a minor inconvenience to realize that the bulk formulae that the model effectively used is different to the one that is described in the ocean model code, it is much more serious to realize that the ocean has received more heat than the atmosphere thought that it gave the ocean. This is a thermodynamic inconsistency that concerns the core property (air-sea heat fluxes) of coupled modelling. Surely, we should not be complacent about allowing some of this heat flux to just disappear.

Fortunately, as discussed in our manuscript, this issue is solved by interpreting the prognostic temperature variable in EOS-80 based ocean models as being Conservative Temperature. We show that doing so means that the equation of state in the EOS-80 based ocean model is not as accurate as it could be, but there are also many other aspects of ocean models that we know are not perfect; just think about the selection of diffusion coefficients, and the temporal drift of deep ocean temperatures. These are all aspects of our science that we oceanographers continue to work on, but at least let's not continue to lose heat at the air-sea interface, especially when we have hit upon such an easy fix.

Finally, potential temperature is no more measurable than is Conservative Temperature. Both are the result of a thought experiment involving an adiabatic and isentropic change of pressure. What is measured is in situ temperature, not potential temperature or Conservative Temperature.

2. $\theta - \Theta$ is a physically-meaningless object that is completely devoid of physical meaning.
Because standard physics teaches us that two quantities can only be compared if they are of the same nature, it follows that one should regard the quantity $\theta - \Theta$ as completely devoid of physical meaning, since it is the difference between a temperature and an energy, whose value depends on the specification of three arbitrary parameters. Yet, the authors seem to suggest that the values of $\theta - \Theta$ - a physically meaningless object - should be regarded as somehow representative of the errors arising from treating θ as conservative in EOS-80 models. How is that possible? Have the authors somehow being taught differently in their physics classes?

First, both potential temperature and Conservative Temperature have the same units (K).

Second, our answer to the referee's Major Point #1 has addressed the issue of the arbitrary constants.

Third, the differences when running an ocean model with potential temperature versus with Conservative Temperature have been derived two ways to date in the literature, as described in the blue text on page 5 above. The difference, $\theta - \Theta$, is representative of the errors made in treating potential temperature as a conservative variable in the case where an ocean model is forced with given surface fluxes [see section A.13 of the TEOS-10 Manual, IOC SCOR and IAPSO (2010)]. In this case there is no error in meridional heat fluxes calculated in either case; the error shows up only in the temperatures themselves [and, as the present manuscript emphasizes, it is all in the *interpretation* of the model's prognostic temperature variable]. With different boundary conditions (such as restoring boundary conditions) the error in assuming that potential temperature is conservative is split in different proportions, between (a) the potential temperature values and (b) the potential temperature fluxes.

3. **One of the premises of the syllogism used by the authors to prove their argument is flawed**. As far as I can judge, the authors arrive at their conclusions that θ should be interpreted as CT in EOS-80 based models by using the following syllogism:

- *Numerical models assume potential temperature to be conservative.*
- *We know that potential temperature is not conservative.*
- *Therefore, potential temperature in EOS80-based ocean models cannot truly be potential temperature and hence should be interpreted as Conservative Temperature.*

Although I agree that the use of syllogisms represents a valid tool in logic to derive a conclusion deductively, it is also well understood that the validity of doing so crucially depends on the validity of the premises. While the validity of `We know that potential temperature is not conservative' is indubitable thanks to McDougall (2003), this is not so of "Numerical models assume potential temperature to be conservative", which is arguably quite a misleading way to characterize pre-TEOS10 ocean modelling practice. Indeed, a much fairer characterization closer to the truth is 'EOS80 based models assume that the errors made in using a constant heat capacity to compute surface fluxes of potential temperature and in neglecting interior non-conservation terms are sufficiently small that they are irrelevant in practice'. If this characterization is used instead of the authors' premise, their syllogism no longer makes sense. As argued above, the authors would have a much stronger case if they could demonstrate the impact of correcting the potential temperature equation in an EOS-80 based model, which would be much more easily understandable by numerical ocean modellers.

The three bullet points at the beginning of this reviewer's 3rd major point do not offer an accurate summary of the thesis of our paper. Rather, the main reason for suggesting that the temperature variable in EOS-80 based coupled models is best interpreted as being CT is that this is the only way to have conservative exchanges of heat between the ocean and atmosphere. Otherwise, coupled climate models lose heat at the air-sea interface, and this is not a good start to the science of climate simulations. That is, we consider it a very basic feature of a coupled climate model that the heat that the atmosphere delivers to the ocean should actually arrive in the ocean (rather than disappearing). Our manuscript shows that this goal is quite easy to achieve.

Our paper examines the pros and cons of requiring that the same heat flux enter the ocean as leaves the atmosphere. That is, we set ourselves the task of enquiring whether it is possible to fix the leaky air-sea interface that EOS-80 based ocean models have when their model temperature is interpreted as being potential temperature. We do this by interpreting the EOS-80 based ocean model's temperature variable as being CT. Much of our paper is about this point, as concisely summarized in Section 6 of the paper.

Also, we do not make any statement about the intentions or assumptions that ocean modellers have when setting up their model equations. Rather, we make statements about what ocean models do. In particular, all ocean models to date, including the EOS-80 based ocean models, treat their model's temperature variable as a conservative variable, and all have a constant value for the isobaric heat capacity.

4. **CT is only 2 or 3 times more conservative than θ.** The conservativeness of CT has been greatly exaggerated so far I admit to being quite confused about the authors' explanation of the origin for the non-conservative production/destruction of θ and CT, which conflicts with the exact results of Tailleux (2010). Indeed, according to the latter, the non-conservative terms $\dot{\theta}_{irr} = \dot{\theta}_{irr}^{diff} + \dot{\theta}_{irr}^{visc}$ and $\dot{\Theta}_{irr} = \dot{\Theta}_{irr}^{diff} + \dot{\Theta}_{irr}^{visc}$ are the sum of a diffusive and viscous contribution, with $\dot{\theta}_{irr}^{visc} \approx \dot{\Theta}_{irr}^{visc} \approx \varepsilon_K$, where ε_K is the local viscous dissipation rate. Yet, the authors's comparison of the relative non-conservativeness of θ and CT appears to be based solely on the diffusive part. Assuming that viscous dissipation balances about 3 TW of mechanical energy input by the wind, tides and surface buoyancy fluxes (which is most likely a significant underestimate) yields an equivalent surface flux of $10 \, mK.m^{-2}$. Based on Graham and McDougall (2013)'s estimates, a summary of the diffusive and viscous contribution to the non-conservation of each variable is therefore:

Non-conservation	diffusive	viscous	sum
$\dot{\theta}_{irr}$	$16 mK.m^{-2}$	$10 mK.m^{-2}$	$26 mK.m^{-2}$
$\dot{\Theta}_{irr}$	$0.3 mK.m^{-2}$	$10 mK.m^{-2}$	$10.3 mK.m^{-2}$

While these results show that Θ is about 50 times more conservative under the action of diffusive mixing -- admittedly an important result in itself and worth mentioning -- its total degree of non-conservativeness accounting for viscous dissipation is only a factor 2-3 better than that of θ overall. A priori, it is the total degree of non-conservativeness that should be compared, not just that due to diffusion effects, in order to establish whether it is justified to treat CT as exactly conservative in a numerical ocean model. Moreover, as shown by Tailleux (2015), the non-conservation arising from using a constant c_p in the estimation of the surface fluxes for θ tends to be balanced by the non-conservation due to diffusive effects, at least globally, which means that one should expect to compensate, at least to some extent. As a result, it is by no means obvious that Prof. McDougall is correct. Again, this could easily be tested by running an EOS-80 based model with a corrected evolution equation for θ.

As a community of scientists, we are not only concerned with accurately modelling the *globally integrated* ocean heat content, but we also want to minimize *regional* errors in heat content and other variables. While the mean error in the global heat content of the ocean may be equivalent to only $10 \, mWm^{-2}$ or $16 \, mWm^{-2}$ (depending on which study one selects), the regional variation of the depth-integrated non-conservative diffusive source term of θ is ten times this, at $\pm 120 \, mWm^{-2}$ (see Graham and McDougall (2013) and Table 1 of our manuscript). This range of $240 \, mWm^{-2}$ contains 95% of the values of the depth-integrated diffusive non-conservative source terms in the θ budget. This error is much larger than the contribution of the viscous dissipation to the global heat budget; this is true no matter whether 1TW or 3TW is assumed for this number.

Looking forward to a future generation of ocean models, we note significant research effort being directed towards carrying a prognostic equation for kinetic energy and its dissipation ε. When this research matures, it will be possible to add ε into the temperature evolution equation as the ocean model is run forward in time. Doing so with Θ as the model temperature variable makes sense (since the diffusive non-conservative error in Θ is no more than 1mK, see Table 1) but makes no sense if θ were the model's temperature variable. This is because the errors we have identified of $\pm 120 \, mWm^{-2}$ in the missing depth-integrated diffusive source terms of θ are much larger than those arising from by ignoring the dissipation of kinetic energy ε. So we disagree with the reviewer's reasoning.

Specific Comments

1. Line 63: *namely heat content ("temperature")* Given that standard thermodynamics teaches us that the concepts of "heat" and "temperature" should not be confused, starting the paper by confusing the two concepts does not bode well for the following, especially coming from the previous chair of WG127 and current chair of JSC who are supposed to teach us the right way of doing thermodynamics.

We agree that these two are different - this is the reason why "temperature" was in quote marks. Since we then spent 5 manuscript pages explaining this difference in section 2 "Background: thermodynamic measures of heat content" we clearly agree with the reviewer and we certainly are not trying to mislead. In response to this concern, we have reworded to

" …heat content (or its related parameter, temperature)…"

2. Lines 64-65. *For computational reasons, it is useful for numerical schemes involved to be conservative [...]* The authors seem to make this a central tenet of their argumentation, even though adding non-conservative terms in a conservative equation does not pose any particular challenge from a numerical viewpoint. Why do the authors consider it would problematic to add the missing non-conservative terms in the potential temperature equation and use the correct boundary condition for the surface fluxes of θ ?

One answer to this question of the reviewer is that we are dealing with existing models and model runs, and are trying to avoid telling people that their work is useless and must be re-run with a new and untested set of non-conservative equations. The utility of conservative schemes is not a proposition, rather, it is a simple statement of existing practice. Ocean models have, to date, treated both their salinity and their temperature variables as being conservative variables, as manifest by their values being affected by the convergence of fluxes. Doing so enables ocean modellers, and those analyzing the output of ocean models, to check that the numerics are not creating spurious internal sources and sinks, and to check that the particular air-sea flux and geothermal flux that they have been told that the ocean is using, is actually the version that is used in the model. These are very valuable consistency checks that are routinely employed in ocean models used for climate.

We agree that there is no fundamental numerical issue with adding a non-conservative source term to an ocean model code (as we understand it, atmospheric models do it as a matter of course, where their temperature variable is sometimes in situ temperature). However, doing so would mean abandoning this valuable checking procedure for the ocean. The more important issue, as explained above in the blue text on pages 3-4, is that it is not possible to accurately deduce the source terms to add to an ocean model when its temperature variable is taken to be potential temperature; it is neither possible at the sea surface nor in the interior. Nor is it possible to accurately calculate the section-integrated heat flux when the model's temperature variable is potential temperature, because of the complications caused by the for interior source terms.

We contrast this difficulty of potential temperature to the preferred use of Conservative Temperature Θ as the model's temperature variable. Using Θ allows a modeller to retain the conservative numerical transport schemes, allowing for consistency with the actual evolution equations.

In addition, if or when the ocean modelling community wishes to use the modeled dissipation of kinetic energy ε as a source term in the model's temperature evolution equation, this is a sensible thing to attempt when using Θ, but it makes no sense to do so

when using θ as the model's temperature variable because the missing diffusive non-conservative source terms are an order of magnitude larger than the ε source term.

3. Lines 70-72. The property of any "heat" variable to be non-conservative is a generic property of any fluid, which is not limited to seawater, and which can be defined independently of the development of any thermodynamic standard for seawater.

We agree with this point. That is, neither entropy, potential temperature nor Conservative Temperature are 100% conservative. We have stated such in the paper.

4. Lines 75-77. I agree that it is now widely recognized that potential temperature is not truly conservative. However, there is nothing in thermodynamics that says that the appropriate measure of heat should be conservative or approximately conservative, which it seems important to point out. The idea that "heat" should be conservative is idiosyncratic to Prof. McDougall and has absolutely no root in classical thermodynamics or anywhere else in the development of the subject. The only two conservative quantities for seawater idealized as a binary fluid are salinity and total energy. As shown by Tailleux (2010), it is not possible for total energy and any `heat' variable to be simultaneously conservative. Assuming `heat' to be conservative is strictly equivalent to assuming that total energy is not conservative, which the authors appear to have overlooked. Given that recent developments seem to focus on the construction of energetically consistent models, e.g., Eden et al. (2014), one should anticipate that ocean modellers will seek to understand how to add the missing non-conservation of "heat" in their models in order to achieve total energy conservation.

The reviewer is incorrect to say that "The idea that "heat" should be conservative is idiosyncratic to Prof. McDougall …". Rather, in McDougall's publications he has always identified all the non-conservative source terms of Θ, including the contribution of the dissipation of turbulent kinetic energy ε, to the non-conservative production of Θ. And when ocean models become sufficiently sophisticated that ε is available to the model at run time, then the largest source of the non-conservation of Θ will be able to be incorporated. When or if ε is incorporated into a model's temperature equation, there will, however, be an operational question/trade-off decision to make between;
   (i)   making the evolution equation of Θ even more accurate by including ε, or,
   (ii) retaining the conservative nature of the temperature variable so as to continue to enable overall volume-integrated tracer budget numerical checks, and to ensure basin-wide heat budgets can still be performed without the complications caused by accounting for interior source terms.

Note also, as described above, this incorporation of ε into a model's temperature equation will not be able to be done accurately by retaining potential temperature as the model's temperature variable.

A significant part of this comment relates to a proposed alternative, Total Energy. As we explain in the appendix to this reply, Total Energy is in fact not conservative, nor is it a "potential" variable, nor (since it includes kinetic and gravitational potential energy) is it purely thermophysical, making it difficult to work with in a practical sense.

5. Lines 70-80. *This empirical fact is an inherent property of seawater*. I disagree. Nearly all fluids a priori suffer from the same problem, as can easily be demonstrated.

We are puzzled here. We say that potential temperature is produced or destroyed by mixing in seawater, and RT disagrees by saying that this is true for nearly all fluids. We did not make any claims about fluids other than seawater, so it seems we are in agreement.

6. Lines 82-83. *[...] results in inherent contradictions*. "Contradiction" is a loaded word here, because all that the authors have established so far is that treating potential temperature and Conservative Temperature as conservative is only approximate, and that the approximation is a better one for the latter than for the former. Using the term "contradiction" frames the problem as one that should be solved by logic alone, whereby an illogical approach can only be corrected by a logical one. In contrast, "an approximation" can be improved by using a more accurate formulation, such as would be achieved by adding the missing non-conservation terms in the potential temperature equation and using the correct flux of potential temperature. In using the term "contradiction", the authors signal their intention of rooting their arguments in a "right" versus "wrong", rather than by a direct demonstration based on comparing two EOS-80 based model using the incorrect and correct evolution equation for θ .

We agree and have thus removed "inherent" to not mislead the reader into thinking there is a logical fallacy. We kept "contradiction" since, even for Θ, there are ignored source terms. As the reviewer correctly notes, our favoring Θ is due to its substantially smaller source term and thus its substantially smaller "contradiction".

7. Lines 88-89. *even at the cost of introducing problems elsewhere* Why would we want to [do] that when solutions exist to solve problems without adding new ones elsewhere?

Our paper is aimed at those ocean model groups that have not yet changed to using TEOS-10. We provide them a way to interpret their model output in a way that has the heat flux that the atmosphere thinks that it delivers to the ocean, to actually arrive in the ocean (rather than disappearing). We regard this to be an essential property of a coupled atmosphere-ocean-ice model, particularly when the aim is to simulate climate, past, present and future. Further, we refer the reviewer again to the issues with implementing non-conservative source terms for potential temperature listed in the blue text on pages 3-4 above.

8. Lines 94-96. *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 global warming*   This is quite a misleading statement, given that these errors are at least partly compensated by the error arising from neglecting the non-conservation of potential temperature, as suggested by the results of Tailleux (2015).

We disagree. What we say here is correct and is supported by our Table 1. The paper of Tailleux (2015) only addresses the *volume-integrated* heat budget, in which there is the well-known (since 2003) significant cancellation between positive and negative source terms of potential temperature. We are concerned with more than simply the volume-integrated heat budget. Graham and McDougall (2013) shed some light on the *regional* variation of the air-sea flux errors if potential temperature is used as the model temperature variable. They exhibited these errors in the form of depth-integrated heat imbalances as a function of longitude and latitude.

9. Lines 96-99. Heat is not a conservative property but total energy is. Why do the authors insist on conserving heat but not total energy? Why do they consider it is more logical or rational to conserve heat but not total energy?

See the appendix to this reply. Total Energy is not a conservative variable.

10. Lines 105-108. *It is well known that in-situ temperature is not an appropriate measure of the "heat content" [...]* I find it very strange that the authors should discuss what is or what is not an appropriate measure of heat content in the absence of consensus on what should be the "true" definition of heat content.

We have changed the wording to "it is well-known that in-situ temperature is not a satisfactory measure of the "heat content" of a water parcel…"

11. Lines 119 Section A.17 of IOC et al. (2010). It is interesting to see that this Appendix only discusses the diffusive part of the non-conservation of θ and Θ, completely overlooking the role of viscous dissipation, and that only the non-rigorous derivations of Graham and McDougall (2013) are cited when Tailleux (2010) gives the exact and explicit forms of non-conservation for the Navier-Stokes equations written in terms of both θ and Θ.

Tailleux (2010) was not the first paper to derive the evolution equations for potential temperature, entropy and Conservative Temperature. Rather it was McDougall (2003) that did this first for all three variables. Furthermore, (i) Tailleux (2010) is based on the incorrect assumption that Total Energy is a conservative variable (again, see the Appendix to this Reply to the reviewer), and (ii) McDougall (2003) and Graham and McDougall (2013) derive their equations for turbulent mixing in the ocean, whereas the equations developed by Tailleux (2010) and Tailleux (2015) apply to molecular diffusion. As shown by Graham and McDougall (2013), when one replaces the molecular fluxes with turbulent fluxes, the correct turbulent results of Graham and McDougall (2013) are not recovered. These issues are fully explained in the Appendix to this Reply to the reviewer.

12. Line 141. Why is it unfortunate?

It would be fortunate if such a heat-like variable was available since it would mean that the present practice of ocean models of treating their temperature variable as being conservative would be exact rather than approximate.

13. Line 149-150 *[...] has a mean non-conservation error in the global ocean of only about* $0.3 \, \mathrm{mW \, m^{-2}}$. As shown by Tailleux (2010) (his Equation (25)) and Tailleux (2015) (his equation (26)), the exact expression for the non-conservative production/ destruction of Θ is

$$\dot{\Theta}_{irr} = \frac{1}{c_p^0} \left[\varepsilon_k - \mathbf{F}_S \cdot \nabla \left(\mu - \frac{T\mu_r}{\theta} \right) - \mathbf{F}_\Theta \cdot \nabla \left(\frac{Tc_p^0}{\theta} \right) \right] \tag{1}$$

and is seen to include the viscous dissipation term ε_k, which the authors subsequently say is of the order of $3mW.m^{-2}$. Presumably, the $0.3mW/m^2$ only refer to the 'diffusive' nonconservation of Θ. The true nonconservation of Θ is therefore at least $3.3mW.m^{-2}$, which is only 5 times less than the $16mW.m^{-2}$ mean nonconservation of θ. This number increases to $10.3mW.m^{-2}$ if a more realistic estimate of $10mW.m^{-2}$ is used for viscous dissipation, as mentioned in my major comments section.

Please see our response to Specific Comment 14 immediately below.

14. Re-reading Graham and McDougall (2013), I realise that the authors estimated the rate of viscous dissipation ε_k from the formula $\varepsilon_k = DN^2/\Gamma$. Their estimate corresponds to a surface integrated value of $3.10^{-3} \times 3.10^{14} = 0.9 \, \mathrm{TW}$, which is way too small. Clearly, there is at least 3 TW of mechanical energy input due to the wind and surface buoyancy fluxes, not more. A more reasonable estimate of the total viscous dissipation is therefore closer to $10 \, \mathrm{mW.m^{-2}}$, which is only a factor of 2 smaller than the nonconservation of θ.

We are not only concerned with modelling the *globally integrated* ocean heat content correctly, but we also wish to minimize *regional* errors in heat content and other variables. While the error in the global heat content of the ocean associated with interpreting the model's temperature variables as being θ may be equivalent to only $10 \mathrm{mWm^{-2}}$ or $16 \, \mathrm{mWm^{-2}}$ (depending on which study one selects), the regional variation in the depth-integrated non-conservative diffusive source terms in the θ budget is ten times this, at $\pm 120 \mathrm{mWm^{-2}}$ (see Graham and McDougall (2013) and Table 1 of our manuscript). This error is much larger than the contribution of the viscous dissipation to the global heat budget, no matter whether 1TW or 3TW is assumed for this number.

15. Lines 163-164. Not really, given the above arguments. The authors are clearly applying double standards here.

Please see our response to Specific Comment 14 immediately above.

16. Lines 165-167. This is a revisionist view of history, given that Conservative Temperature was introduced by McDougall (2003), long before the IAPWS group was formed, and is not actually part of the UNESCO endorsed part of TEOS-10 (as far as I am aware). This decision was Prof. McDougall's alone, and was not part of the TEOS-10 work. The SCOR/IAPSO WG127 was approved in 2005, and its first meeting took place in Warnumünde in May 2006.

We thank RT for emphasizing the need for historical accuracy.

Conservative Temperature, CT, is part of TEOS-10. When TEOS-10 was adopted by the IOC, the TEOS-10 Manual became the official description of TEOS-10, and in the TEOS-10 Manual there are many sections that derive the evolution equations and recommend the use of Conservative Temperature for analysing ocean observations and in ocean models.

The Chairman of the IOC, the President of SCOR and the President of IAPSO announced the replacement of EOS-80 with TEOS-10 in a series of scientific journals including the following announcements in Ocean Modelling and Deep-Sea Research,

https://www.sciencedirect.com/science/article/pii/S1463500311001545

or https://doi.org/10.1016/S1463-5003(11)00154-5

and

https://www.sciencedirect.com/science/article/pii/S0967063711001348

or https://doi.org/10.1016/j.dsr.2011.07.005

In these announcements it says, among other things, "In particular, Conservative Temperature Θ accurately represents the ''heat content'' per unit mass of seawater, and is to be used in place of potential temperature θ in oceanography."

We agree that the idea of Conservative Temperature predates WG-127. The history of Conservative Temperature in TEOS-10 can be read in the following article on the history of the development of TEOS-10 in this paper on which Remi Tailleux is a co-author (see particularly pages 167-170):

https://os.copernicus.org/articles/8/161/2012/ (Pawlowicz, R., T. McDougall, R. Feistel and R. Tailleux, 2012: An historical perspective on the development of the Thermodynamic Equation of Seawater – 2010: *Ocean Sci.*, **8**, 161-174. http://www.ocean-sci.net/8/161/2012/os-8-161-2012.pdf)

In brief, the idea that potential enthalpy might be a good heat-like variable came to TMcD on Cape Cod while swimming in Crooked Pond before breakfast on 1st July 1994. By morning tea that day he was convinced that the idea was viable. It was presented to Nick Fofonoff in the form of a seminar to him and his students a few weeks later. Nick had no objections to the idea, but he must have then forgotten about it as he thereafter co-authored the Bacon and Fofonoff (1996) paper. The idea did not appear in the literature until nine years after 1994, in McDougall (2003), although another paper [Cunningham, S. A., 2000: Journal of Marine Research, 58, 1-35.] did use the idea, quoting a pre-print of what became McDougall (2003).

Perhaps we should have said "officially define" instead of "define" – since any paper can define something but TEOS-10 has an official standing. We have thus replaced "to define a new variable," with "to adopt …".

17. Lines 180-183. The question of why potential temperature is non-conservative was actually answered earlier by Tailleux (2010), who showed that the non-conservation of any heat variable is dictated by the first law of thermodynamics (the law of total energy conservation). This is again a revisionist view of history where Graham and McDougall (2013) attempts to get credit for something that needs to be attributed to Tailleux (2010).

We disagree. The original development of the non-conservation of potential temperature, entropy and in particular, Conservative Temperature, was published by McDougall (2003), not Tailleux (2010). Moreover, as explained in the Appendix to this reply, the Tailleux (2010) paper claims to be based on the assumption that the Total Energy is a conservative variable. This is incorrect; Total Energy is not a conservative variable. However, in fact, Tailleux (2010) ignored the key non-conservative production term, $-\nabla \cdot (P\mathbf{u})$, and actually arrived at correct expressions for the non-conservative production of potential temperature, Conservative Temperature and of entropy (Eqn. (B.7) of Tailleux (2010) and Eqn. (B10) of Graham and McDougall (2013) are identical). But these expressions are written in terms of the molecular fluxes of heat and salt; fluxes that include the Soret and Dufour effects. In their Appendix B, Graham and McDougall (2013) showed that the expressions of Tailleux (2010), being written in terms of the molecular fluxes of heat and salt, cannot be used by simply replacing the molecular fluxes with turbulent ones; doing so leads to disobeying the

Second Law of Thermodynamics.

By contrast, McDougall (2003) and Graham and McDougall (2013) derived their key results by considering turbulent mixing in the ocean. Their equations do obey the Second Law of Thermodynamics, and because (apart from double-diffusive effects) mixing processes in the ocean are turbulent, the approach of these papers is applicable to the ocean, and to ocean models.

18. Lines 199 and below. How conservative is conservative temperature? As said above, this section only describes the non-conservation of Conservative Temperature arising from diffusive mixing and completely omits viscous dissipation. Given that the latter dominates, this section is at best misleading.

We have now taken the opportunity to repeat a reference, at this place in the text, to the fact that ε also contributes to the non-conservation of Θ. This is also the place in our manuscript where we have inserted reference to the erroneous assumption made by both Tailleux (2010) and Tailleux (2015) that Total Energy $\mathcal{E}$ is conserved when fluid parcels mix.

19. Line 382. What the authors call naive is what I call common sense. The authors should expect that many oceanographers will feel insulted here.

We do not aim to insult. We have thus replaced "naïve" with "At first glance then, it seems reasonable…"

20. Line 448. I dispute that this is thermodynamic best practice if the authors fail to understand the results of Tailleux (2010). .

We hope that the Appendix to this reply makes it clear that the authors do understand the results of Tailleux (2010) and Tailleux (2015), but that that they disagree with the foundation of these papers. Both of these papers stated that Total Energy $\mathcal{E}$ is a conservative quantity (which it is not), but then these papers ignored the non-conservative production term, $-\nabla\cdot\left(P\mathbf{u}\right)$, that appears in the evolution equation of $\mathcal{E}$. By making these two self-balancing errors, Tailleux (2010) arrived at correct expressions for the non-conservative production of potential temperature, Conservative Temperature and of entropy, but these expressions were written in terms of the molecular fluxes of heat and salt, and since the mixing processes at work in the ocean are turbulent fluxes, these expressions are not immediately applicable. This is explained in more detail in our Appendix to this Reply.

21. Lines 479. The authors confuse the terms "contradiction" and "approximation" -- This is an idiosyncratic interpretation because this is not how idealized modelling should be viewed. Numerical modellers and oceanographers understand that potential temperature is non-conservative; it is therefore unfair to accuse them of assuming potential temperature to be conservative. Rather, they treat it as conservative because they assume that the small non-conservative terms and heat flux errors do not matter on the time scales generally considered. "Contradiction" and "approximation" are two completely different concepts, which it is crucial to distinguish in the present discussion, since the authors use the first interpretation in order to be able to accuse EOS-80 based modelling practice as being illogical. The interpretation "Models

assume the potential temperature to be conservative" has been disproven and there is no reason to accuse ocean modellers of ignoring this result. On the other hand, the interpretation that "models assume that the non-conservation of potential temperature is sufficiently small that it can be neglected in practice" has not been disproven yet. Indeed, disproving such an approach can only be achieved by running an EOS-80 based model with the incorrect and corrected equation for potential temperature. Such experiments are urgently needed so that we can stop with all the speculation.

We appreciate the need to distinguish "contradiction" (a logical fallacy) from "approximation". Our usage of "contradiction" is clearly in reference to the errors associated with the interpretation of the model's temperature variable as potential temperature rather than Conservative Temperature, with "contradiction" referring to the inaccuracies that are now well documented in the literature. Furthermore, there are numerical modellers co-authoring this paper, and they have no concern with pointing out these inconsistencies. The modelling community must move forward. Furthermore, as explained in our response to RT's comments on lines 64-65, the proposal from Tailleux (2015) that potential temperature should be retained as the temperature variable of an ocean model, together with adding source terms, is unworkable, because it cannot be done accurately.

22. Lines 735. It is not true that Conservative Temperature is consistent with the first law of thermodynamics, because it assumes that all the heat goes into heat and none into work. Indeed, it is well known from Lorenz's theory of available potential energy that there is about 0.5TW of the surface buoyancy fluxes going into the production of available potential energy. This suggests that potential enthalpy includes APE -- a work-like quantity -- as well as heat, not just heat.

This point has been well discussed in the literature, namely that changing from potential temperature to Θ reduces the non-conservative diffusive source terms' contribution to heat fluxes (additional to the ε term) by a factor of ~100. Importantly, in adopting Θ in place of potential temperature, there has been no claim that there is zero production of available potential energy. We do not say or imply that.

McDougall (2003) and Graham and McDougall (2013) avoid these pesky issues surrounding Total Energy, available potential energy and the like. Rather than being concerned with variables such as these Total-Energy-and-its-constituent energy components, McDougall (2003) and Graham and McDougall (2013) consider quite a different variable, namely the potential enthalpy h^m referenced to the pressure p^m at which an individual mixing event takes place. The pursuit of h^m has proved rewarding ever since 1st July 1994.

Summary of our Reply to the Reviewer

The main disagreements that this reviewer has with our paper are based on the following three viewpoints with which we disagree,

    (i)   a concentration by the reviewer only on the terms that contribute to the globally volume-integrated heat content of the ocean, rather than also being concerned with errors in modelling temperature in specific geographic locations,

    (ii)   a hope (that also pervades the Tailleux (2015) paper), that retaining potential temperature as a model's temperature variable could be made to be competitive, in terms of accuracy and practicality, compared with adopting Conservative Temperature in numerical ocean circulation models, and,

    (iii)   what we believe is an error made in Tailleux (2010) and Tailleux (2015) in the choice of physical property that is assumed to be conserved under turbulent mixing.

The error made by Tailleux (2010) and Tailleux (2015), referred to in point (iii) above, underlie many of this reviewer's comments; in particular, his "Specific comments" numbered 4, 9, 10, 11, 17, 20 and 22. This error is the assumption that Total Energy, $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$, is a conservative variable. Because this issue underlies so many of the reviewer's comments and is central to both the Tailleux (2010) and Tailleux (2015) papers, TMcD directly addresses this issue in the following Appendix rather than in our direct response to the review.

The research of McDougall (2003), Graham and McDougall (2013), Tailleux (2010) and Tailleux (2015) have all followed a similar path in that they worked from a variable that was taken to be conserved when two fluid parcels mixed in the ocean, and then they used this assumed knowledge to derive expressions for the non-conservation of Θ, θ and η. While the approach of Tailleux (2010, 2015) is flawed because Total Energy was taken to be to be conservative, these papers also ignored the key non-conservative term, $-\nabla\cdot\left(P\mathbf{u}\right)$, so that they actually arrived at the correct evolution equations for Θ, θ and η (for example, Eqn. (B.7) of Tailleux (2010) and Eqn. (B10) of Graham and McDougall (2013) are identical). However, these equations are written in terms of the molecular fluxes of heat and salt, and it is not possible to use these expressions to evaluate the non-conservation of these variables for turbulent mixing processes. If this is attempted, an erroneous term of magnitude $0.2\rho\varepsilon/\hat{h}_{\Theta}$ appears.

Before beginning the Appendix, we authors admit that two of us have given lectures and drafted other material that has also stated that Total Energy is a conservative variable (which it is not). These statements have since been corrected.

Appendix: Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$ is not a conservative variable

24th April 2021 by Trevor J McDougall

In this appendix we prove that Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$ is not a conservative variable. As a reminder, a variable C is called "conservative" if, when two fluid parcels are turbulently mixed together, the total amount of C-substance in the mixed fluid is the sum of the amount of C-stuff in the two initial fluid parcels.

We note that the papers of McDougall (2003) and Graham and McDougall (2013), and the approach taken by TEOS-10 have based their analyses upon

(i) the conservation of potential enthalpy referenced to the pressure of a mixing event,
(ii) rather than [as advocated in Tailleux's review of our manuscript, and by Tailleux (2010) and Tailleux (2015)] upon the assumed conservation Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$.

The evolution equations of several energy-type variables

Here the evolution equation equations are listed for several variables, copied from the TEOS-10 Manual, IOC et al. (2010), (but here without the terms due to the non-conservation of Absolute Salinity).

$$\left(\rho\left[\tfrac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi\right]\right)_t + \nabla\cdot\left(\rho\mathbf{u}\left[\tfrac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi\right]\right)$$
$$= \rho\,\mathrm{d}\left(\tfrac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi\right)/\mathrm{d}t = -\mathbf{u}\cdot\nabla P + \nabla\cdot\left(\rho v^{\mathrm{visc}}\nabla\tfrac{1}{2}[\mathbf{u}\cdot\mathbf{u}]\right) - \rho\varepsilon. \tag{B.12}$$

$$\rho\left(\frac{\mathrm{d}h}{\mathrm{d}t} - \frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}t}\right) = \rho\left(\frac{\mathrm{d}u}{\mathrm{d}t} + P\frac{\mathrm{d}v}{\mathrm{d}t}\right) = \rho\left((T_0+t)\frac{\mathrm{d}\eta}{\mathrm{d}t} + \mu\frac{\mathrm{d}S_\mathrm{A}}{\mathrm{d}t}\right)$$
$$= -\nabla\cdot\mathbf{F}^\mathrm{R} - \nabla\cdot\mathbf{F}^\mathrm{Q} + \rho\varepsilon. \tag{B.19}$$

$$\left(\rho\mathcal{E}\right)_t + \nabla\cdot\left(\rho\mathbf{u}\mathcal{E}\right) = \rho\,\mathrm{d}\mathcal{E}/\mathrm{d}t = -\nabla\cdot\left(P\mathbf{u}\right) - \nabla\cdot\mathbf{F}^\mathrm{R} - \nabla\cdot\mathbf{F}^\mathrm{Q} + \nabla\cdot\left(\rho v^{\mathrm{visc}}\nabla\tfrac{1}{2}[\mathbf{u}\cdot\mathbf{u}]\right). \tag{B.15}$$

$$\left(\rho\mathcal{B}\right)_t + \nabla\cdot\left(\rho\mathbf{u}\mathcal{B}\right) = \rho\,\mathrm{d}\mathcal{B}/\mathrm{d}t = \partial P/\partial t\big|_{x,y,z} - \nabla\cdot\mathbf{F}^\mathrm{R} - \nabla\cdot\mathbf{F}^\mathrm{Q} + \nabla\cdot\left(\rho v^{\mathrm{visc}}\nabla\tfrac{1}{2}[\mathbf{u}\cdot\mathbf{u}]\right). \tag{B.17}$$

$$\partial\left(\rho u\right)/\partial t + \nabla\cdot\left(\rho\mathbf{u}u\right) = \rho\frac{\mathrm{d}u}{\mathrm{d}t} = -P\frac{1}{v}\frac{\mathrm{d}v}{\mathrm{d}t} - \nabla\cdot\mathbf{F}^\mathrm{R} - \nabla\cdot\mathbf{F}^\mathrm{Q} + \rho\varepsilon. \tag{B.18}$$

$$\partial\left(\rho h\right)/\partial t + \nabla\cdot\left(\rho\mathbf{u}h\right) = \rho\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{\mathrm{d}P}{\mathrm{d}t} - \nabla\cdot\mathbf{F}^\mathrm{R} - \nabla\cdot\mathbf{F}^\mathrm{Q} + \rho\varepsilon. \tag{A.13.2}$$

Eqn. (B.12) is the evolution equation for the sum of kinetic and gravitational potential energies. Eqn. (B.19) is the First Law of Thermodynamics (with its first line being the Fundamental Thermodynamic Relationship), while Eqns. (B.15) and (B.17) are, respectively, the evolution equations for Total Energy, $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$, and of the Bernoulli function, $\mathcal{B} = h + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$. Eqn. (B18) is the evolution equation for internal energy, u, and Eqn. (A.13.2) is the evolution equation for enthalpy, h. In these equations the source term due to the non-conservative nature of Absolute Salinity has been omitted, for simplicity. Specific entropy is η. The radiative flux of heat is $\mathbf{F}^\mathrm{R}$, while $\mathbf{F}^\mathrm{Q}$ stands for the molecular heat flux and the air-sea and geothermal fluxes of enthalpy, while ε is the rate of dissipation of kinetic energy.

What have McDougall (2003), Graham & McDougall (2013) and TEOS-10 done?

The benefits of potential enthalpy and Conservative Temperature have largely been justified in these papers of 2003, 2013, and in the TEOS-10 Manual, from the viewpoint of evolution equations, but the benefits, and the original motivation that led to McDougall (2003), can be summarized by the following physically motivated three-point argument.

1.   First, note that the air-sea heat flux is a flux of potential enthalpy, h^0, namely the flux of potential enthalpy referenced to a fixed surface reference pressure [and h^0 is, by decree, exactly c_p^0 times the flux of Conservative Temperature].

2.   Second, note that the Taylor-series analysis of the non-conservative production of potential enthalpy, h^0, (as can be found in appendix A.18 of the TEOS-10 Manual) shows that while it is the potential enthalpy h^m referenced to the pressure p^m of a mixing event, that is conserved when parcels mix at p^m, a negligible error is made when Θ is assumed to be conserved during the mixing event at pressure p^m. This estimating of the non-conservation of h^0 proceeds from the conservation of h^m is described in detail in Graham and McDougall (2013) and is summarised in Appendix A.18 of the TEOS-10 Manual. Most of this small production of Θ is due to the dissipation of kinetic energy, ε, and a much smaller part is due to the inherent (or diffusive) non-conservation of Θ. The conservation equation for h^m, for a mixing process occurring at pressure p^m is

$$\partial\left(\rho h^m\right)\big/\partial t + \nabla\cdot\left(\rho \mathbf{u} h^m\right) \;=\; \rho\frac{\mathrm{d}h^m}{\mathrm{d}t} \;=\; -\nabla\cdot\mathbf{F}^{\mathrm{R}} - \nabla\cdot\mathbf{F}^{\mathrm{Q}} + \rho\varepsilon. \qquad (\text{A.13.2 at } p^m)$$

3.   Third, note that the ocean circulation can be decomposed into a series of adiabatic and isohaline movements during which Θ is absolutely unchanged (because of its "potential" nature), followed by a series of turbulent mixing events during which Θ is almost totally conserved (see point 2).

These three points, taken together, show that Θ is the quantity that is advected and diffused in an almost conservative fashion and whose surface flux is exactly proportional to the air-sea heat flux.

Point 2 above relies on the conservative nature of enthalpy when mixing occurs at a given pressure, p^m. This conservative nature can be deduced by examining Eqn. (A.13.2). At this pressure p^m, the left-hand side of Eqn. (A.13.2) is the evolution equation of potential enthalpy, h^m, referenced to the pressure p^m, while the right-hand side contains (1) the divergence of molecular and radiative heat fluxes and (2) the positive-definite dissipation term $\rho\varepsilon$. Integrating over a material volume that is larger than, but contains the two parcels that are undergoing mixing (even when the volume changes shape and the volume decreases during the mixing process), we deduce that the non-conservation of potential enthalpy h^m is caused <u>only by</u> the dissipation term $\rho\varepsilon$. The use of h^m rather than enthalpy itself allows fluid parcels from say 1 dbar above p^m and 2 dbar below p^m to be moved adiabatically and isentropically to the pressure at which the mixing event occurs, while exhibiting no change in h^m during these motions that occur prior to the mixing event itself.

Before moving on, this volume-integrated conservation statement will be derived in detail. When Eqn. (A.13.2 at p^m) is spatially and temporally integrated over a moving and contracting volume in which a mixing event occurs, the Leibnitz differentiation of the volume integral ensures that the relevant velocity that appears at the surface of the volume is the velocity *through* this moving boundary, the dia-surface velocity, $\mathbf{u}^{\mathrm{dia}}$. This can be understood by

considering the time differentiation of the volume integral of the total amount of h^m-substance in the volume, as on the left-hand side of the first line of the equation below. Now, doing this differentiation of the volume integral with respect to time, the last term on the right-hand side of the first line of this equation arises from the fact that the boundary is moving through space, with $\mathbf{u}^{\text{boundary}}$ being the velocity of the bounding surface of the volume. In the second line, Eqn. (A.13.2 at p^m) has been used to replace the temporal derivative $\partial(\rho h^m)/\partial t$ term that appears in the first line. The third line converts four of the volume integrals into boundary area integrals using the divergence theorem (and $\mathbf{u}^{\text{dia}} = \mathbf{u} - \mathbf{u}^{\text{boundary}}$).

$$
\begin{aligned}
\frac{\partial}{\partial t}\left(\int_V \rho h^m \, dV \right) &= \int_V \left(\rho h^m\right)_t dV \; + \int_S \rho h^m \mathbf{u}^{\text{boundary}} \cdot d\mathbf{S} \\
&= -\int_V \nabla\cdot\left(\rho h^m \mathbf{u} + \mathbf{F}^R + \mathbf{F}^Q\right) dV \; + \int_S \rho h^m \mathbf{u}^{\text{boundary}} \cdot d\mathbf{S} + \int_V \rho\varepsilon \, dV \\
&= -\int_S \rho h^m \mathbf{u}^{\text{dia}} \cdot d\mathbf{S} - \int_S \left(\mathbf{F}^R + \mathbf{F}^Q\right)\cdot d\mathbf{S} \; + \int_V \rho\varepsilon \, dV
\end{aligned}
$$

The volume we chose to examine is a material volume whose boundary moves with the fluid parcels. Since at the edge of this volume there is no mixing, the fluid and the volume move together and there is no dia-surface velocity, so the first term on the right-hand side of the last line is zero. Since there are no radiative or molecular fluxes of heat at the boundary of the volume, the middle terms on the last line are also zero. This leaves just the last term on this line of the equation. Hence we see that the difference between the final and the initial volume integrals of ρh^m, that is, the difference in the total amount of h^m- substance in this volume after the mixing event compared with before the mixing event, is the time integral of the volume integral of $\rho\varepsilon$ over the mixing event. No other energy-like variable is as clean as this.

In summary, McDougall (2003) and Graham and McDougall (2013), as well as the TEOS-10 recommendations, are based on the almost-conservative nature of $h^m = h\left(S_A, \Theta, p^m\right)$, for turbulent mixing process that occurs at pressure p^m. In contrast, Tailleux (2010, 2015) and the many remarks of this review of our paper, have assumed that a different variable, namely Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$, is an almost-conserved variable when fluid parcels mix at p^m. We will now show that this is not the case. Moreover, the TEOS-10 approach has yielded expressions for the non-conservation of potential temperature, Conservative Temperature and entropy that apply to an ocean in which the mixing processes are turbulent, whereas the Tailleux (2010, 2015) papers have led to expressions for the non-conservation of these quantities that apply only when the mixing in the ocean is done by molecular diffusion.

What did Tailleux (2010) and Tailleux (2015) do?

Background

The papers Tailleux (2010), Tailleux (2015), and the review of our paper by Remi Tailleux, all take Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$ to be a conservative variable. Here, following the work of McDougall, Church and Jackett (2003),

McDougall, T. J., J. A. Church and D. R. Jackett, 2003: Does the nonlinearity of the equation of state impose an upper bound on the buoyancy frequency? *Journal of Marine Research*, **61**, 745-764. http://dx.doi.org/10.1357/002224003322981138

we will show that $\mathcal{E}$ is not a conservative variable. Actually, after having assumed that $\mathcal{E}$ is a conservative variable, the papers Tailleux (2010, 2015) then went on to ignore the non-conservative source term $-\nabla\cdot(P\mathbf{u})$ that is part of the evolution equation for $\mathcal{E}$, and so these papers ended up deriving correct expressions for the non-conservative production of potential temperature, Conservative Temperature and of entropy. However these expressions were written in terms of the molecular fluxes of heat and salt, and, as shown in Appendix B of Graham and McDougall (2013), these expressions cannot be used by simply replacing the molecular fluxes with turbulent ones. Hence these expressions are not able to be used in the ocean or in ocean models where the mixing processes are predominantly turbulent.

What about the $-\nabla\cdot(P\mathbf{u})$ term when mixing occurs?

The advantage of the evolution equation for the Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$, Eqn. (B.15), is that its right-hand side is the divergence of a flux. This is by construction; this is how the Total Energy evolution equation is derived. Given the divergence nature of the right-hand side of Eqn. (B.15), there is a tendency to assume that this means that Total Energy is conserved when two fluid parcels mix. This is not the case, as will now be demonstrated.

Following McDougall, Church and Jackett (2003), consider the mixing of two seawater parcels at pressure p^m in a single vertical water column (say, in a tall rigid tube of constant area, as illustrated in Figure 4 of that paper, repeated below). After the parcels have mixed, the final volume of the mixed fluid is less than the sum of the two initial volumes (note that this non-conservative decrease in volume on mixing is a general property of seawater and many other, although not all, fluid mixtures), no matter what the initial contrasts in temperature and salinity, see Graham and McDougall (2013)). This contraction-on-mixing at p^m means that due to this mixing, the total ocean volume has decreased, with seawater parcels at shallower locations slumping to slightly greater depths. When considering all these fluid parcels that lie above the mixing event, the $-\nabla\cdot(P\mathbf{u})$ source term can be written as the non-zero term $-\nabla\cdot(P\mathbf{u}) = -\mathbf{u}\cdot\nabla P = \partial P/\partial t\big|_{x,y,z}$ since all the fluid parcels move in an adiabatic, isohaline and isobaric manner with $\nabla\cdot\mathbf{u} = 0$ and with $\mathrm{d}P/\mathrm{d}t = 0$. These parcels, which were not involved in the mixing at pressure p^m, retain their Absolute Salinity, their Conservative Temperature, their enthalpy, their internal energy, and their pressure, but their height, z, and their geopotential, $\Phi = gz$, has decreased. These fluid parcels thus experience a **decrease** in their Total Energy, that is, $\mathrm{d}\mathcal{E}/\mathrm{d}t < 0$, even though they have experienced no mixing.

Now consider what occurs at the location of the mixing at p^m. While the average height and geopotential of this fluid has changed a little during the mixing, the larger contribution to the change in Total Energy $\mathcal{E}$ is due to the work done by the surrounding ocean, via the pressure acting on the change in volume, $-P\nabla\cdot\mathbf{u}$. With the Absolute Pressure (in Pa) at the mixing location being P^m, the **increase** in the Total Energy of this mixed fluid arising from the

$-\nabla\cdot\left(P\mathbf{u}\right)$ term on the right-hand side of Eqn. (B.15) is approximately P^m times the volume-and-time integral of $-\nabla\cdot\mathbf{u}$ which, by using the continuity equation, can be seen to be the volume-and-time integral of $-v^{-1}\mathrm{d}v/\mathrm{d}t$ over the water being mixed. As shown by McDougall, Church and Jackett (2003), the gravitational potential energy that all the fluid above the mixing location loses (adiabatically and isentropically) due to contraction-on-mixing during the mixing process at p^m, is close to that gained by the mixed fluid in this non-conservative fashion as an increase of the mixed fluid's internal energy, u. The difference between these quantities is the work done by the atmospheric pressure acting on the changing height of the sea surface due to the contraction-on-mixing, since this adds to the depth-integrated Total Energy of the water column.
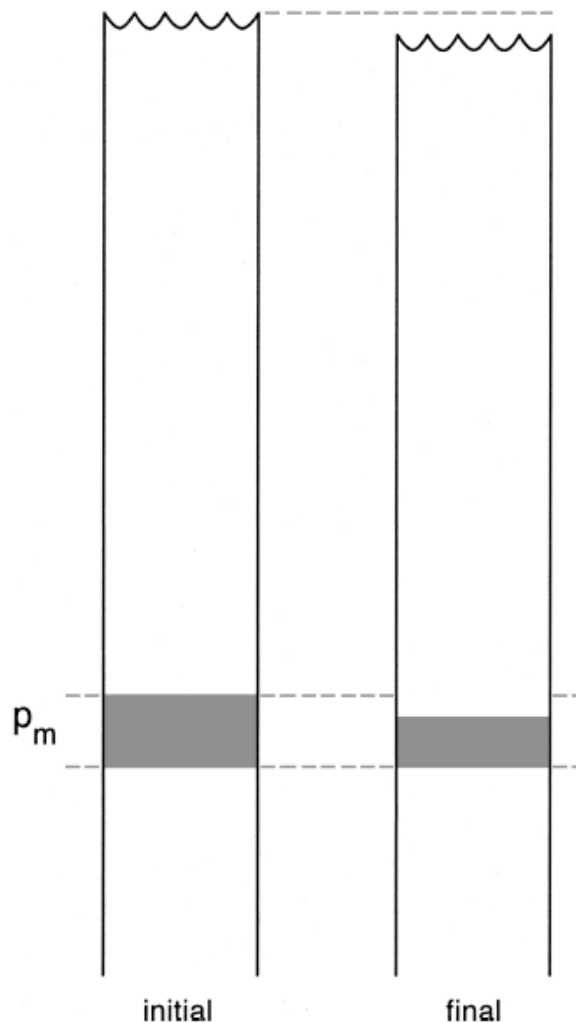

Figure 4. Sketch of the slumping of a strictly one-dimensional water column (like a very tall test tube) due to turbulent mixing in the vicinity of pressure p_m. Each parcel in the water column above p_m retains its enthalpy, its internal energy and its pressure. The reduction in gravitational potential energy of the entire water column appears as an increase in the internal energy of the fluid that undergoes the mixing.

There is a tendency to think that because the $-\nabla\cdot\left(P\mathbf{u}\right)$ term is the divergence of a flux, that it will integrate to zero over a patch of turbulent mixing, however this is not the case. The reason is that at the boundary of a fixed volume in space in which the mixing is occurring, the velocity of the fluid is non-zero, and part of this is due to the contraction-on-mixing. In the evolution equation (A.13.2 at p^m) the divergence term $\nabla\cdot\left(\rho\mathbf{u}h^m\right)$ can be integrated to zero over the mixing volume because of the presence of the temporal term, $\partial\left(\rho h^m\right)/\partial t$, in that equation (this was explained in detail on page 22 above). By contrast, in the case of the

$-\nabla\cdot\left(P\mathbf{u}\right)=-\nabla\cdot\left(Pv\,\rho\mathbf{u}\right)$ term, there is no corresponding temporal term, $\partial\left(\rho\,Pv\right)/\partial t\,=\,\partial P/\partial t$, in Eqn. (B.15) with which to form a paired Leibnitz differentiation of the volume integral. In fact, by adding and subtracting this temporal derivative term, one arrives at the evolution equation for the Bernoulli function, Eqn. (B.17).

To be totally explicit about the non-conservation of Total Energy, the analysis that is presented on page 22 above, is now performed for Total Energy, $\mathcal{E}\,=\,u+\tfrac{1}{2}\mathbf{u}\cdot\mathbf{u}+\Phi$.

$$
\begin{aligned}
\frac{\partial}{\partial t}\left(\int_V \rho\mathcal{E}\,\mathrm{d}V\right) &= \int_V \left(\rho\mathcal{E}\right)_t\,\mathrm{d}V \;+\; \int_S \rho\mathcal{E}\,\mathbf{u}^{\text{boundary}}\cdot\mathrm{d}\mathbf{S} \\
&= -\int_V \nabla\cdot\left(\rho\mathcal{E}\,\mathbf{u}+P\mathbf{u}+\mathbf{F}^{\mathrm{R}}+\mathbf{F}^{\mathrm{Q}}-\rho v^{\text{visc}}\nabla\tfrac{1}{2}\left[\mathbf{u}\cdot\mathbf{u}\right]\right)\mathrm{d}V \;+\; \int_S \rho\mathcal{E}\,\mathbf{u}^{\text{boundary}}\cdot\mathrm{d}\mathbf{S} \\
&= -\int_S \rho\mathcal{E}\,\mathbf{u}^{\text{dia}}\cdot\mathrm{d}\mathbf{S} - \int_S\left(\mathbf{F}^{\mathrm{R}}+\mathbf{F}^{\mathrm{Q}}-\rho v^{\text{visc}}\nabla\tfrac{1}{2}\left[\mathbf{u}\cdot\mathbf{u}\right]\right)\cdot\mathrm{d}\mathbf{S} - \int_S P\mathbf{u}\cdot\mathrm{d}\mathbf{S}
\end{aligned}
$$

Again, the right-hand side of the first line is the Leibnitz temporal differentiation of the volume integral on the left, the second line has used the evolution equation (B.15) for Total Energy, while the third line has collected two of the advection terms into one, involving the dia-surface velocity $\mathbf{u}^{\text{dia}}=\mathbf{u}-\mathbf{u}^{\text{boundary}}$. For our turbulent mixing scenario of the above figure, the first and middle surface integrals on the last line of this equation are both zero, but the last term, being the volume integral of $-\nabla\cdot\left(P\mathbf{u}\right)$ is non-zero, and because of this non-zero term, Total Energy $\mathcal{E}$ is not a conservative variable.

Total Energy is not a "potential" variable

This thought process has shown that Total Energy is not conserved when two fluid parcels mix at a given pressure. Neither is Total Energy a "potential" variable; this can be seen as follows. Noting that the total energy $\mathcal{E}$ is related to the Bernoulli function by $\mathcal{E}=\mathcal{B}-Pv$ and even if we take the whole ocean to be in a steady state so that $\mathcal{B}$ has the "potential" property (see Eqn. (B.17)), it is clear that $\mathcal{E}$ does not have the "potential" property in this situation. That is, $\mathcal{E}$ is not "quasi-material". In such a steady-state ocean, the change in $\mathcal{E}=\mathcal{B}-Pv$ caused by an adiabatic and isohaline change in pressure of 1 Pa can be shown to be $-v\left(1-Pv/c^2\right)$ where c is the sound speed. This expression is only slightly different to $-v$, so that for an increase of pressure of $1000\,\text{dbar}$, the decrease in total energy $\mathcal{E}$ is approximately the same as that caused by a decrease in Conservative Temperature of $\sim 2.4°\text{C}$. This means that, even in a steady-state ocean, total energy $\mathcal{E}$ is as useless as is enthalpy h as far as being a marker of fluid flow, since they are both very far from being "potential" variables.

Total Energy is not a thermophysical variable

Neither is Total Energy $\mathcal{E}$ a thermophysical variable since it is not a function only of Absolute Salinity, temperature and pressure.

$-\nabla\cdot\left(P\mathbf{u}\right)$ *is not a "reversible" mixing term*

The approach taken by all of the above papers, namely McDougall (2003), Graham and McDougall (2013), Tailleux (2010) and Tailleux (2015), is to start with a variable that is taken to be conserved (under certain assumptions), when seawater parcels mix at constant pressure p^m. This assumed knowledge is then used to find the expressions for the non-conservative production of Conservative Temperature, potential temperature, and of entropy. So, at the core of this technique is to begin with a variable that is conserved when fluid parcels mix at pressure p^m.

When potential enthalpy $h^m = h\left(S_A, \Theta, p^m\right)$ is the variable-of-choice, the only non-conservation is due to the sign-definite dissipation of kinetic energy, ε; the last term in Eqn. (A.13.2). When Total Energy $\mathcal{E} = u + \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + \Phi$ is taken to be the variable of choice, the non-conservation occurs even when there is no mixing taking place locally. Moreover, concentrating now on the location where the mixing is occurring, the non-conservative production of Total Energy is proportional to the absolute pressure, P^m, and from Figure 3(c) of McDougall, Church and Jackett (2003) we see that the non-conservative production of Total Energy $\mathcal{E}$ is larger than the dissipation of kinetic energy ε in 1.8% of the volume of the global ocean.

However, in a further wrinkle, despite claiming that Total Energy is a conservative variable, Tailleux (2010) and Tailleux (2015) do not actually take the right-hand side of the Total Energy evolution Eqn. (B.15) equation to be zero. Rather, these papers assume that mixing does not contribute to the $-\nabla\cdot\left(P\mathbf{u}\right)$ term, and its influence is not examined. While this term is clearly the divergence of a flux, we have shown that this term cannot be ignored in the quest to quantify the non-conservation of variables such as Θ, θ and η.

Rather, as we have shown above, the $-\nabla\cdot\left(P\mathbf{u}\right) = -P\nabla\cdot\mathbf{u} - \mathbf{u}\cdot\nabla P$ term is non-zero when mixing occurs in the ocean; the first term on the right here, $-P\nabla\cdot\mathbf{u}$, causes a non-conservative increase in $\mathcal{E}$ at the location of the mixing, while the second term, $-\mathbf{u}\cdot\nabla P$, causes $\mathcal{E}$ to not be conserved when a water column loses gravitational potential energy, even when this occurs by an adiabatic and isentropic slumping motion, due, for example, to mixing occurring deeper in the water column.

Evolution equations for Θ, θ and η in terms of the molecular fluxes of heat and salt

The evolution equation for Θ, written in terms of the molecular fluxes of heat and salt, is

$$\rho\frac{d\Theta}{dt} = -\nabla\cdot\left(\frac{\mathbf{F}^R + \mathbf{F}^Q - \hat{h}_{S_A}\mathbf{F}^S}{\hat{h}_\Theta}\right) + \left(\mathbf{F}^R + \mathbf{F}^Q\right)\cdot\nabla\left(\frac{1}{\hat{h}_\Theta}\right) - \mathbf{F}^S\cdot\nabla\left(\frac{\hat{h}_{S_A}}{\hat{h}_\Theta}\right) + \frac{\rho\varepsilon}{\hat{h}_\Theta} \qquad \text{(CT molecular1)}$$

(Tailleux (2010) and Eqn. (B4) of Appendix B of Graham and McDougall (2013)) where the partial derivatives $\hat{h}_\Theta$ and $\hat{h}_{S_A}$ of $\hat{h}\left(S_A, \Theta, p\right)$ are functions of S_A, Θ and pressure. This equation can be found by simply rearranging the First Law of Thermodynamics, Eqn. (A.13.2), after taking enthalpy to be in the functional form $\hat{h}\left(S_A, \Theta, p\right)$. The molecular fluxes of salt and heat in this expression are given by the following expressions from the TEOS-10 Manual,

$$\mathbf{F}^S = -\rho k^S\left(\nabla S_A + \frac{\mu_P}{\mu_{S_A}}\nabla P\right) - \left(\frac{\rho k^S T}{\mu_{S_A}}\left(\frac{\mu}{T}\right)_T + \frac{B}{T^2}\right)\nabla T , \qquad (B.26)$$

$$\mathbf{F}^Q = -\frac{1}{T^2}\left(C - \frac{B^2}{A}\right)\nabla T + \frac{B\mu_{S_A}}{\rho k^S T}\mathbf{F}^S = -\rho c_p k^T\nabla T + \frac{B\mu_{S_A}}{\rho k^S T}\mathbf{F}^S, \qquad (B.27)$$

which involve the molecular diffusivities of salt, k^S, and temperature, k^T, as well as the cross-diffusion coefficient B. These expressions contain the well-known cross-diffusion Soret and Dufour fluxes, as well as the flux of salt down the gradient of pressure.

Is this Θ evolution equation, Eqn. (CT molecular1), useful in quantifying the non-conservation of Θ in the ocean and in ocean models, given that we know that the dominant mixing processes are turbulent, not molecular? During the turbulent mixing process the average of the middle terms in this equation, $\left(\mathbf{F}^R + \mathbf{F}^Q\right)\cdot\nabla\left(1/\hat{h}_\Theta\right) - \mathbf{F}^S\cdot\nabla\left(\hat{h}_{S_A}/\hat{h}_\Theta\right)$, using Eqns. (B.26) and (B.27) for $\mathbf{F}^S$ and $\mathbf{F}^Q$, involve complicated products of the gradient of in situ

temperature, the gradient of pressure, the gradient of Conservative Temperature, and the gradient of Absolute Salinity. These products of spatial gradients then need to be averaged over the time and space of the turbulent mixing event. This formidable set of correlations have not yet proved possible to understand. Eqn. (CT molecular1) can also be written in the form

$$\rho \frac{d\Theta}{dt} = -\nabla \cdot \mathbf{F}^\Theta - \frac{\mathbf{F}^\Theta \cdot \nabla \hat{h}_\Theta}{\hat{h}_\Theta} - \frac{\mathbf{F}^S \cdot \nabla \hat{h}_{S_A}}{\hat{h}_\Theta} + \frac{\rho \varepsilon}{\hat{h}_\Theta}$$ (CT molecular2)

where the molecular flux of Conservative Temperature $\mathbf{F}^\Theta$ can be shown to be $\left(\mathbf{F}^R + \mathbf{F}^Q - \hat{h}_{S_A} \mathbf{F}^S \right) / \hat{h}_\Theta$, but this does not simplify the problem of averaging over all the spatial gradients that are involved in the expressions for the molecular fluxes. We do know, from Appendix B of Graham and McDougall (2013), that it is not possible to simply replace the molecular fluxes of CT and salt in Eqn. (CT molecular2) with the corresponding turbulent fluxes. If this is attempted, an erroneous term of magnitude $0.2 \rho \varepsilon / \hat{h}_\Theta$ appears. Hence the laminar evolution equation of Conservative Temperature in which the fluxes are molecular fluxes, as derived in Tailleux (2010), does not lead to a means of accurately quantifying the non-conservation of Conservative Temperature in the turbulently mixed ocean.

What other heat-like variables have been proposed in the literature?

Prior to Conservative Temperature being adopted by the oceanographic community, the dominant heat-like variable whose net meridional flux in the ocean was compared to the corresponding air-sea heat flux, was potential temperature multiplied by a fixed isobaric heat capacity.

Bacon and Fofonoff (1996) [Bacon, S. and N Fofonoff, 1996: Oceanic heat flux calculation. *J Atmos. Oceanic Technol.* **13**, 1327-1329] advocated a different measure of heat content for oceanographic use, namely, $c_p\left(S_A, \theta, p^0\right)\theta$, being potential temperature multiplied by the isobaric heat capacity that the seawater parcel would have if moved adiabatically and isentropically to the sea surface pressure. Section 7 of McDougall (2003) analysed this proposal and showed that this variable is no more accurate as a measure of the heat content of a fluid parcel than is θ.

Warren (1999) [Warren, B. A., 1999: Approximating the energy transport across oceanic sections. *J. Geophys., Res.*, **104**, 7915-7919.] proposed the use of internal energy, u, but, as explained above, internal energy is not conserved when fluid parcels mix, with the non-conservation being due to the work done by the environment's pressure as the volume reduces. Warren (1999) also proposed $\overline{c_p}\theta$ as an approximation to internal energy, where $\overline{c_p}$ is the average value of the isobaric heat capacity evaluated at the sample's salinity, the sea surface pressure, and over a range of potential temperatures between zero Celsius and the parcel's potential temperature θ. McDougall (2003) showed that this variable was not a particularly good approximation to internal energy, but that it was in fact equal to $h^0\left(S_A, \theta\right) - h^0\left(S_A, 0\right)$, that is, to the potential enthalpy of the fluid parcel minus the potential enthalpy of a fluid parcel at the same salinity but at zero Celsius temperature; clearly, this second term is unwanted.

This short discussion illustrates that various authors have searched for a heat-like variable whose transport in the ocean can be accurately compared with the air-sea heat flux.

Summary of the Appendix

This appendix can be summarized as follows. Graham and McDougall (2013) have demonstrated that potential enthalpy h^m referenced to the pressure p^m at which a mixing event occurs, is

(i)   a "potential" variable,

(ii)   a thermophysical variable, and

(iii)   an "isobaric conservative variable" for turbulent diffusive mixing processes at pressure p^m; the dissipation of kinetic energy can be added as a source term.

In contrast, Total Energy $\mathcal{E}$ has none of these three desirable attributes.

Tailleux (2010, 2015) claimed that Total Energy is a conservative oceanographic variable, but this is not the case; rather, there is the non-conservative source term $-\nabla \cdot \left(P \mathbf{u} \right)$ in the evolution equation for $\mathcal{E}$. Tailleux (2010, 2015) then went on to ignore this source term and actually arrived at the correct set of evolution equations for potential temperature, Conservative Temperature, and entropy, but these evolution equations were written in terms of the molecular fluxes of heat and salt, and are not applicable to the turbulently mixed ocean. So, to date, only the approach of McDougall (2003) and Graham and McDougall (2013) has yielded the evolution equation for Conservative Temperature that is applicable to a turbulent ocean.