Review of Whitwell et al. (2026) submitted to Ocean Science

Title: Observations of the multi-year variability of mixing in shelf seas (egusphere-2026-2498).

This manuscript presents 5 year long timeseries of temperature measurements from four different mooring sites on the northwestern Australian shelf. It analyses temporal variability in both the diapycnal diffusivity and diapycnal heat flux using observed temperature data from the moorings. The manuscript considers the impact of tidal and seasonal variability on the variability of the diffusivity and heat flux. A final analysis considers the impact of the different sampling / deployment periods on the accuracy of the estimated K_ρ and J_Q.

The considered dataset is an interesting and extensive dataset and it is a relevant topic for publication in Ocean Science. However, at the moment the manuscript is a lot of show and tell, while deeper analysis is often missing. Furthermore, I do have some serious concerns regarding the used methods and the corresponding results. Also the conclusions are relative descriptive and offer little discussion and critical reflection. In its current form, the conclusions would benefit from additional supporting information and clearer justification. More critically connecting the findings to the physical processes and existing literature would substantially improve the impact and value of this manuscript.

A more detailed description of my concerns, as well as some minor comments and technical points that will help with improving the clarity of the text and figures are listed below. Given these concerns, I recommend major revisions are made before publication.

Major comments

1. I have concerns about the validity of using the Ellison scales in the (surface) mixed layer. In the mixed layer (or any weakly stratified environment for that matter), the Ellison scales can be undefined. In a shallow shelf-sea environment the mixed layer can span a significant portion of the watercolumn (as is evident from Fig. 3) and can thus potentially seriously affect the results. Did you test or validate in any way whether using Ellison scales is a valid approach here? Does it affect the results? It would be good to see some comparison with diffusivity estimates from other studies and methods, e.g. from microstructure observations etc. Also it would be appreciated if some reflection on this could be added to the Discussion section.

2. Research question 2 in the Introduction states that one of the aims is to see whether the diapycnal diffusivity or vertical heat flux is a better diagnostic for the variability of the mixing. However, l. 41-42 does not provide an explanation why the heat flux is necessary. I.e., this does not explain in what ways the 'normal' method of the diffusivity would be inadequate, or what would be gained from using the heat flux instead. Please elaborate on what criteria you judge either of the variables to perform better or to be more adequate.

   Then, the Abstract and Conclusions claim that the vertical heat flux performs better than the diffusivity. Yet, little supporting evidence is provided of this claim. Please explain where and why the heat flux performs better in the presented results. E.g. what aspects were not covered by the diffusivity? Also, for the BAR, ROW and BRW sites the vertical structure of K_ρ and J_Q is similar, however, for the LYN site this appears not to be the case. Any ideas why there is this difference?

   Lastly, in the current form K_ρ and J_Q are not independent metrics, where the diffusivity is directly used for the calculation of the heat flux. For the comparison it would be much stronger if diffusivity estimates obtained from e.g. shear measurements would be used. I understand that this is likely not an option, given instrumental and operational constraints. But some discussion on this, and the effect it has on the conclusions would be good.

3. It is unclear from the manuscript whether there are specific reasons for not adopting the TEOS-10 framework. Given that TEOS-10 is the current standard in physical oceanography, the authors should clarify which thermodynamic formulation is being used and provide justification if they intentionally chose not to use TEOS-10 variables and derived quantities. There may be valid methodological or operational reasons for this choice regarding the derived quantities such as density

and buoyancy frequency, but these are not presently discussed. In any case, I expect the authors to use the TEOS-10 variables and the correct notation for the observed variables.

Minor comments / Technical points

l. 7: site

l.18: I dont see the intermittency as the main problem. First and foremost it is about the required sensor sensitivity.

l.20: Also here the intermittency is a secondary problem. In the first place, models are unable to resolve the scales of mixing at all.

l. 29-32: The context from literature could be more extensive. I get that these timeseries are scarce, but none of those works is currently being cited (e.g. Moum et al., 2009, 2023; Van Haren et al., 2024; to name a few, but there are plenty others as well). Also it would be good to make the distinction between timeseries from direct microstructure measurements, or from ADV's, and timeseries from indirect methods, such as is the case here. Furthermore, are moorings the only platforms collecting timeseries or are other platforms, e.g. gliders, also being used?

l. 38 e.g. instead of i.e.

l.39: It is more appropriate to refer to the actual models here. I.e. Osborn (1980); Osborn, and Cox (1972).

l. 72: The geographical extend of the mooring sites does not provide a lot of information. For someone not familiar with the region, please provide some more context on the climate and atmospheric forcing. What kind of climate is prevalent? Tropical, temperate etc? The conclusions mention tropical cyclones, do these occur often? And are the conditions for the mooring sites the same? (This comment also applies to various places throughout the manuscript that reference to discussion about atmospheric forcing, but none is provided. Some references that I spotted are: l. 2, 179, 221, 235, 350, but there might be more)

l.73: (And all other figures in general) I can see why you chose for this panel numbering in the figures, but it makes for a bit of awkward figure referencing. It would be neater to label figure panels with incrementing letters only.

l. 103: This suggests that there was more instrumentation on the moorings than only the ADCP and thermistors, like SBE37 Microcat CTD's or similar? Please provide details of the instrumentation used.

l.115: Reference missing.

Sec. 2.3 (and other sections throughout the document): A lot of sentences start with 'We': 'We characterized... . We calculated... .' It is distracting from the contents and it breaks the flow of the text. I would advise some careful rereading and in some places rephrasing of the text.

Figure 2. I find this figure unclear and a bit confusing. From the text I take there was some standard design of the mooring. Of course there will some variations because of redeployments. I would remove Fig 2 as it is, and replace it with a sketch / overview of the standard mooring. This also allows for showing additional instrumentation, e.g. CTD's.

l.146: It seems that a different dt/dz threshold is being used for defining weak stratification. It would make sense to use the same threshold throughout.

l. 152: Equation (1) (Please also check this for all other equation references)

l. 165: (Fig. 3d)

l. 166: Any ideas what can be causing this surprising stratification pattern?

Figure 3. As the main datasource of this manuscript is the thermistor data. It would be helpful to show the recorded temperature data, similar as is done for N in Figure 3. This can be done next to the panels of Figure 3, or as a separate figure. Please also add the observed mixed layer depth to the panels of both variables.

Did you consider using a logarithmic colorscale for N? It might be more appropriate, though it is hard to judge without knowing the full range of the data.

Furthermore, I am more than interested in seeing the full timeseries of K_ρ and J_Q, similar to the presentation as N. Given the first part of this comment, I probably would add the temperature record to Fig 3, next to the timeseries of N, and a separate figure with the timeseries of K_ρ and J_Q.

l. 181-187, Fig. 5: This is a relevant analysis, but it currently does not contribute to the interpretation of the mixing variability at the mooring sites. You can either remove this part and the figure entirely, or, connect it more with the existing literature (it very well connects with the descriptions in Sec 2.1) and describe how it can be seen in the observed variables.

Figure 5: comma missing after 'tidal axis'

l. 192: Reference to Fig 6

l. 199: I disagree with this statement. Without the corresponding temperature record it is hard to judge whether the increased heatflux is equivalent to increased mixing. The shown diffusivities suggest that the mixing is not necessarily more active for the ROW and BRW sites.

l. 200: I disagree with this statement. For the BAR, ROW and BRW sites the vertical structure of both K_ρ and J_Q shows similarities. Only for the LYN site there is a notable difference. Please also explain how you see the effect of the stratification in this.

l. 208: Can you offer some hypothesis? E.g. based on the slope criticality? Or the roughness of the topography?

Fig 7: Axis labels and units are missing.

l. 220: The statement about the LYN site is incorrect. The energy levels at the bottom depth are clearly higher than for those at the middle and top depths.

l. 220: There has been no discussion of the wind so far. Either add that discussion somewhere or remove this statement.

l. 221: But the LYN site also had weak tidal forcing, but has similar spectral energy levels as the other sites. Please provide some more analysis and interpretation of the results.

l. 246: I get that because you are using a linear equation of state, N is directly proportional to the temperature gradient. But in essence, the difference between your diffusivity and heat flux is a factor dt/dz. So also for that it would be good to see the temperature record.

l. 251: A Hilbert transform might not be clear to every reader, a reference would be helpful

l. 257-264: The spring-neap variability for the ROW site is not evident to me. The differences in the average K_ρ and J_Q between spring and neap are not any bigger than for the BAR and LYN site. I do see a slight increase in the variability of the bottom two measurements. Please revise your analysis here. Possibly you can link it with the earlier analysis of the spectra.

Sec 3.4: This is a nice section to bring the different analysis together. But at the moment it is mostly a description what can be seen in Fig. 11. Please provide some more analysis. Also link it with the previous figures. E.g the dominance of the summer contribution at the top depth at BAR (Fig. 11, b4) is not directly explained when comparing it to Fig (8, c4). Similar questions can be asked about the switch in spring-neap (Fig. 10, b3) and why the summer bottom flux at LYN is so large (75%), while the difference in Fig (8, c1) between the summer and winter seems small. More analysis and comparison with the previous figures and results would help this section a lot.

l. 298: all play

Fig 11: x-axis label missing

Sec 3.5: This section is, in my opinion, superfluous. It is more or less a repeat of the discussion from Sec. 3.2 and the difference in interpretation between median and average mixing values. Scheifele et al. (2021) (You already cited them in Sec 3.2) describes this nicely: the median value describes the typical mixing rate, or heat flux in this case. Whereas the average value describes the cumulative effect over the entire record. With the 6hr averaging, the averaging is only minimal, and thus this distribution will very much resemble the variability seen in Fig. 6. By increasing the averaging window, it is obvious that the variability will converge to the record average value.

l. 321: '-' missing between 'J_Q' and 'average'

l. 330: space missing between 'J_Q' and 'depending'

Sec 4: In addition to the multiple points raised that call for more discussion and reflection. It would be good to also refer more explicitly back to the research question raised in the Introduction.

References

Moum, J. N., and J. D. Nash, 2009: Mixing Measurements on an Equatorial Ocean Mooring. *Journal of Atmospheric and Oceanic Technology*, **26**, 317–336, 10.1175/2008jtecho617.1.

Moum, J. N., W. D. Smyth, K. G. Hughes, D. Cherian, S. J. Warner, B. Bourlès, P. Brandt, and M. Dengler, 2023: Wind Dependencies of Deep Cycle Turbulence in the Equatorial Cold Tongues. *Journal of Physical Oceanography*, **53**, 1979–1995, 10.1175/JPO-D-22-0203.1.

Osborn, T., 1980: Estimates of the Local Rate of Vertical Diffusion from Dissipation Measurements. *Journal of Physical Oceanography*, **10**, 83–89, 10.1175/1520-0485(1980)010⟨0083:EOTLRO⟩2.0.CO;2.

Osborn, T. R., and C. S. Cox, 1972: Oceanic Fine Structure. *Geophysical Fluid Dynamics*, **3**, 321–345, 10.1080/03091927208236085.

Scheifele, B., S. Waterman, and J. R. Carpenter, 2021: Turbulence and Mixing in the Arctic Ocean's Amundsen Gulf. *Journal of Physical Oceanography*, **51**, 169–186, 10.1175/JPO-D-20-0057.1.

van Haren, H., G. Voet, M. H. Alford, B. Fernández-Castro, A. C. Naveira Garabato, B. L. Wynne-Cattanach, H. Mercier, and M. J. Messias, 2024: Near-slope turbulence in a Rockall canyon. *Deep-Sea Research Part I: Oceanographic Research Papers*, **206**, 10.1016/j.dsr.2024.104277.