Response to reviewer 1

Throughout this document, the two reviews are reproduced in black. Our responses are presented in blue and, where appropriate, quotations from the revised paper are included in indented *italic blue* text.

REVIEWER 01

In this manuscript, Sheehan et al. describe results from a Seaglider mission in January-February 2020 in the western tropical Atlantic. The paper focuses on reporting dissipation rates of turbulent kinetic energy using two methods: a direct microstructure method, with which temperature fluctuations are measured at centimetre-scales, and an indirect method based on the meter-scale vertical overturns detected in the temperature profiles (the so-called Thorpe method), also measured with a microstructure probe. The most interesting pattern in the observations is a local enhancement of turbulent dissipation below a subsurface layer of high-salinity associated with the subtropical underwater. The authors claim that this is the first study reporting microstructure temperature measurements from this particular type of glider. Another highlight of the paper is that, contrary to some previous assessments, the direct and indirect methods for estimating dissipation rates agree well in their dataset. Finally, the authors also report and discuss heat and salt fluxes driven by the measured turbulent motions are favourable in the region.

Although the paper has some strong points (it describes novel dataset with some interesting results, it is well written), it has also – in my view – some major weaknesses that need to be addressed before publication. My general feeling is that the authors missed several good opportunities to make the paper more relevant and useful for their colleagues. I list my comments below with some suggestions of further analysis/discussions that could potentially improve the impact of the article:

We thank the reviewer for recognising the value of the dataset and the work, and for the thorough suggestions they have made regarding its improvement. We agree that they will increase the impact of the article.

(1) Microstructure Method. Overall, I missed a more extensive and direct acknowledgement of the limitations of the different methods used. The authors use a nowadays standard method to estimate energy dissipation rates from temperature microstructure measurements (the Batchelor fitting method). However, from my experience and the available literature, the method has important limitations and the results may be highly dependent on some user choices. In my opinion, those aspects are relevant to discuss. One key limitation is the fact that the dissipation rates detectable with the method are upper bounded due to the limited time response of the FP07 sensor. For typical glider along-path speeds of 40 cm/s, maximum dissipation rates measurable are about 10 W/Kg. The authors report values larger than that in the upper mixing layer, but I haven't seen any mentioning as to how the sensor limitation may affect their results, in particular those values, and the comparison with the Thorpe-scale method. I would like to see some of the temperature-gradient spectra in the upper mixing layer and the fit parameters, to confirm that those values correspond to real dissipation. My experience is that for high

epsilon values the roll-off in the temperature gradient is probably imposed by sensor limitations, such that the Batchelor fitting is not really meaningful (see also Paterson and Fer 2014).

We thank the reviewer for bringing this limitation to our attention. Following Peterson and Fer (2014), we will remove any estimate of epsilon greater than 2×10^{-7} W kg⁻¹. We will add this as a quality control criterion in Section 2.2:

"The estimate of ε_{μ} is greater than 2 × 10-7 W kg-1; even having corrected estimates of Λ_{32} at high frequencies, the effective resolution of the fast thermistor (100~Hz) is such that values of ε_{μ} greater than this cannot be reliably estimated."

Furthermore, we will outline the expected influence of sensor limitation on our results in Section 3.1:

"But these highest values [$\varepsilon_{\mu} > 10^{s}$] are infrequently observed [in the surface mixed layer]: a large proportion of ε_{μ} estimates in the mixed layer are greater than 2×10^{-7} W kg⁻¹, beyond the range for which the FP07 fast thermistor observations and the Batchelor spectrum method yield meaningful results (Sec. 2.2; Peterson and Fer, 2014). Consequently, the remaining values are all below this threshold; indeed, considerable variability in ε_{μ} is observed in the mixed layer, and many remaining estimates are low (< 10⁻¹⁰). The mean ε_{μ} in the upper water column is therefore likely to be biased towards these low values."

The time-depth plot of estimates of ε_{μ} with the addition of the new upper-limit cut off ($\varepsilon_{\mu} > 2 \times 10^{-7}$ W kg⁻¹) are presented in (Figure 1a) at the end of this document. Examples of relatively high ε_{μ} ($10^{-8} < \varepsilon_{\mu} < 2 \times 10^{-7}$ W kg⁻¹) spectra from the mixed layer are presented in (Figure 2) at the end of this document. We note that estimates of ε_{μ} within this range are few, given that many already fail the quality control procedure of Scheifele et al (2018) and Scott et al (2021).

Further, the temperature spectra are typically partially corrected for the sensor time-response. The best parameters for the adjustment are not well stablished and can differ from sensor to sensor. It would be good to report explicitly what exact correction is used here.

We apologise for omitting this information in our summary of the method of Scheifele et al. (2018). Once the temperature power spectrum, Λ_{32} , has been calculated from a 32-second segment, the transfer function of Sommer et al. (2013) is used to correct for the slow FP07 response time at high frequencies. This step is included in the Matlab toolbox provided by Ben Scheifele and Jeffrey Carpenter. We will add this step into our summary of the method.

"Values of Λ_{32} at high frequencies, where the thermal inertial of the fast thermistor is such that its temporal response is inadequate, are corrected using the transfer function of Sommer et al. (2013)."

I also have a general vocabulary concern. The authors make a difference between what they call "spectral" methods for calculating turbulent dissipation, as opposed to non-spectral methods, like the fine-scale and Thorpe methods. In my view this is not entirely correct since, for example, the fine-scale method often involves spectral calculations. In my opinion, what distinguishes the two kind of methods is the scale at which measurements are done. In this regard, I think it would be more correct to use the term "microstructure method" as opposed to "indirect methods" (or something similar). I would recommend to change this throughout the manuscript.

We agree that, on reflection, spectral methods does not accurately and uniquely describe the methods used in this study. In the revised paper, we will use "microstructure" methods/observations to refer to the Batchelor spectrum method applied to 512 Hz FP07 temperature observations; we will use "finescale" methods/observations to refer to observations of lower resolution (e.g. ship and glider CTD) and to the methods commonly used to process them (e.g. Thorpe scale, finescale parameterisation).

(2) Thorpe-scale method. Regarding the Thorpe method, its applicability to non-vertically profiling instruments like gliders has been questioned in the past (e.g. Thorpe et al. 2012), since the method is based on measuring the vertical size of turbulent overturns, which is not entirely achievable with a slant-wise measuring trajectory. Although your results are encouraging, I still think it would be better to acknowledge and discuss the potential impact of this limitation more explicitly.

We thank the reviewer for highlighting this potential limitation of using the Thorpe scale method on glider observations. Thorpe (2012) suggested that internal waves, when present, can lead to false overturn detection and therefore to artificially high estimates of ϵ_T ; Howatt et al. (2021) further suggest that the detection of false overturns is of concern when the slope of internal waves is greater than the glide slope (i.e. angle from the horizontal) of the glider. Following Howatt et al. (2021), we calculate that the mean (± one standard deviation) glide slope for the present mission is 35 ± 5 degrees – i.e. greater than the maximum slope of internal waves (16 degrees). We will add an acknowledgement of this to the revised paper:

"A potential problem with using the Thorpe scale method on Seaglider observations is the nonvertical profile that the Seaglider collects: unlike, for instance, a ship-deployed CTD or a vertical microstructure profiler, a Seaglider follows a slanted trajectory, typically covering a distance of approximately 4 km over a 1000 m dive-climb cycle. Any resultant sampling of horizontal gradients, particularly in the presence of internal waves, could artificially inflate estimates of ε_T due to false overturn detection (Thorpe 2012). However, this is a concern only when the internal wave slope exceeds the slope of the Seaglider's trajectory. Following Howatt et al. (2021), we calculate that the mean (± one standard deviation) of SG620's trajectory slope (from the horizontal) during the Eurec4a deployment was $35 \pm 5^{\circ}$, i.e. greater than the upper limit of the slope of internal waves (16°; Thorpe, 1978). Consequently, false overturn detection due to the Seaglider's sloping trajectory is unlikely to lead to over-estimation of ε_T in this dataset."

Also, the authors contrast their results (in which the microstructure and Thorpe methods agree well), with a previous study (Howatt et al. 2021) in which the Thorpe method tended to overestimate dissipation. They speculate that this difference may be due to the different probe resolution: in the present study a microstructure probe is used, which allows a much better resolution of the smaller overturns, whereas Howatt et al. (2021) use a standard CTD probe. While this is just a mere speculation, the authors have the available data to evaluate this hypothesis: they could apply the Thorpe method to the glider regular CTD data and compare the results with those of the FP07-Thorpe method. I think not doing that is a missed opportunity, since it would be very useful for future studies to know whether the disagreement comes from the different spatial resolution or not.

We thank the reviewer for this suggestion. We feel that the original wording in our discussions did not fully convey the weight of the findings of Howatt et al. (2021), who clearly conclude that the low resolution of a glider CTD sets a minimum resolvable turbulent patch size. In oceanographic conditions where overturns are likely to occur over much smaller scales than the minimum resolvable turbulent patch size, ensuing estimates of ε_T will necessarily be biased high. We agree with this argument. Our original discussion should have made this clearer. Values of LT derived from the FP07 fast thermistor observations (512 Hz) are clearly far smaller than those derived from ordinary Seaglider observations

(0.2 Hz; Figure 3 of this document), indicating that the latter are unable to resolve the majority of overturns in this region. Estimates of ε_T from these ordinary Seaglider observations will therefore be extremely unreliable. Given that Howatt et al. (2021) have already reported this conclusion, we do not want to repeat their work and include figures illustrating this difference, but we will comment on the inability of the ordinary Seaglider observations to accurately resolve L_T.

I have another concern about the method, in particular about what you describe in line 145 regarding the aggregation of overturns until they reach a scale of 2 m. Sorry, I am not familiar with this approach, but it seems a bit weird to me, why do you need to impose a minimal overturning scale if those appear smaller in the data? This could bias high the lower dissipation rates, couldn't it?

We thank the reviewer for bringing this to our attention. The aggregation of overturns is traditionally done during Thorpe scale calculations with relatively low-resolution observations in order to ensure that there are sufficient data points within each turbulent patch (e.g. Howatt et al., 2021). We originally included this step, despite having very high-resolution observations, for consistency with previous work. However, we are re-examining whether this step is indeed necessary when using 512 Hz observations, and whether it makes an appreciable difference to estimates of ε_T . We will include this in the revised paper.

(3) Salt fingers. I have two main comments regarding the double diffusion diffusivity calculations and fluxes. First, I think there is a clear missed opportunity here of using the microstructure data to assess the role of double diffusion more directly. I missed, for example, a figure showing some of the staircases, but more importantly, you could use the thermal variance dissipation rates to directly assess the rate of heat mixing in the salt-finger favourable zone using the Osborn-Cox method. Several authors (e.g. St Laurent et al. 1999; Schmitt et al. 2005) have reported levels of thermal variance dissipation in thermohaline staircases, and heat diffusivities above those predicted by the Osborn model (your eq. 7). In most microstructure studies based on shear probes this information is missing, but you have it and you could compare the results with the flux law estimates.

As discussed immediately below, we will change the method used to identify those regions susceptible to salt fingering; nevertheless, we will happily include a figure illustrating salt fingers if it adds to the arguments of the revised paper. We agree that an estimate of heat diffusivity based on our estimates of χ would make for an interesting comparison with the existing results. We will include such an estimate in the revised paper, likely as an additional panel in Figure 8.

My second comment concerns the latter method. I see that in your Figure 7, salt fingers diffusivities are zero except in the step layers, however it is not clear how you deal with those zeros in the calculation of mean diffusivities and fluxes. If you just neglect them, you may be overestimating the average diffusivities by quite a lot! But this is not explicit. Can you clarify that? That said, I am not entirely sure about the way in which the flux-law formulae should be applied, but my intuitive guess would be that the equations should be applied for the entire region of the vertical profile where the density ratio is favourable for salt fingers and the steps are observed, that is, applied to the background temperature-salinity gradients instead of the gradients in the steps, as you did. Not sure if I am wrong on that, but I invite you to dive a deeper into it and find out what is the most correct way of applying the formulae and report the details (how and where the diffusivities, gradients and fluxes were calculated; how did you deal with zeros.

Having more closely studied the original paper by Radko and Smith (2012), we agree with the reviewer that Equations 10 and 11 should be applied to the entire region of the vertical profile where the density ratio is favourable for salt fingers, rather than just in the gradient layers. Consequently, we will use the

Turner angle to identify those parts of the water column that are susceptible to salt fingering, and will apply the equations to these regions. Further, we agree that it would be more accurate to calculate a time-depth average while including regions in which the fluxes of heat and salt due to salt fingering are zero – i.e. to include these zeros in our averages. We thank the reviewer for making these suggestions.

(4) Discussion of the observed patterns. The main feature of your dissipation dataset is an enhancement in a thick layer below the salinity maximum. In my view, there is also a missed opportunity here of making this result more meaningful and relevant by discussing the possible drivers of this enhancement. Is that maybe due to internal wave interactions with the rapidly changing stratification? Do you have current measurements from the ship that indicate a local enhancement of shear or other data that could give some clue of what is going on? Not sure if further analysis on this is needed for the manuscript (which has a strong methodological component), but I think it could help to make the paper more relevant and interesting.

As the reviewer suggests, we will compare our results with ADCP observations from the ship that passed closest to the glider during the deployment, and will comment on this in the revised paper.

SPECIFIC COMMENTS:

Line 8. Superscript missing in W kg-1

This will be corrected.

Line 21. The Batchelor method applies only to scalars (e.g. temperature), not shear. The use in this sentence is not correct

This sentence will be altered from:

"Estimating the dissipation rate of turbulent kinetic energy, ε , by applying the Batchelor spectrum method (Batchelor, 1959) to high-resolution observations of shear and temperature (e.g. Lueck et al., 2002; Peterson and Fer, 2014; Scheifele et al., 2018) has, historically, required considerable ship time, plus specialist instruments and expertise."

to:

"Estimating the dissipation rate of turbulent kinetic energy, ε , from high-resolution observations of shear and temperature (e.g. Lueck et al., 2002; Peterson and Fer, 2014; Scheifele et al., 2018) has, historically, required considerable ship time, plus specialist instruments and expertise."

Line 25. These authors report 266 profiles in 50 different stations

We will remove this statement.

Line 61, the use of "," and "()" in this sentence, and other locations in the manuscript, seems weird to me

We will make changes to improve punctuation in various locations in the manuscript.

Line 73. Why didn't you use the shear probe data in this paper? It feels also like a potential missed opportunity of a more thorough evaluation of the different methods.

There was a fault with the shear probe and the data are unfortunately unusable. This will be mentioned in the revised paper:

"There was a fault with the shear probe on this deployment and the observations could not be used."

Line 95. This sentence "the rate of destruction of temperature-gradient variance" is not correct: note that chi units are K^2 per second, not (K/m)^2 per second. Temperature of thermal variance is destroyed, not the temperature-gradient variance. Remove "-gradient"

This will be corrected as suggested.

Line 114–115. I am not familiar with this paper but the normalisation of the temperature gradient spectra with epsilon seems a bit unphysical to me, since the spectrum scales with chi not with epsilon. There is a proper way of normalising the temperature variance spectrum (see e.g. Dillon and Caldwell (1980)).

The full procedure developed by Scott et al. (2021) is described in the supplementary material to the paper we cited (available at *agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020JC016861*). We normalise estimates of ε only to perform this quality control step: it is the original, non-normalised values of ε that are subsequently considered in the paper.

Line 123. You could perhaps define the meaning of the term $(epsilon/N)^{1/2}$. Also note the wrong spelling in "Tayor"

This sentence will be altered from:

"If the quantity $U/(\varepsilon_{\mu} / N)^{1/2}$ is less than five, where U is the glider's speed-in-direction-or-travel, indicating that Tayor's frozen turbulence hypothesis is invalid."

to:

"If the quantity $U/(\varepsilon_{\mu} / N)^{1/2}$ is less than five, where U is the glider's speed-in-direction-of-travel and $(\varepsilon_{\mu} / N)^{1/2}$ is an estimate of the turbulent flow velocities (Fer et al., 2014), indicating that Taylor's frozen turbulence hypothesis is invalid."

Line 161 onwards. How was the critical value chosen? It seems a bit arbitrary. Why don't you use the alpha and beta coefficients to directly estimate the contributions of T and S gradients/variability to density stratification/variability? A more detailed description of the results of this quality check would be desirable. How many datapoints have been discarded? Where?

We thank the reviewer for bringing this to our attention. We agree that using alpha and beta coefficients (i.e. the stability ratio) would be more appropriate as a quality control measure. In the revised paper, we will implement a quality control check based on the stability ratio, and we will provide a detailed description of the impact of this on our dataset.

Figure 3 or 4. It could be nice to see a distribution of LT here or elsewhere.

We agree that this would make a nice addition to the paper. A time-depth section of L_T is presented at the end of this document (Figure 4) and will be included in the revised paper.

Figure 4. Why are there no epsilon_T values in the upper mixing layer? What is the criterion to discard those values? (I guess it is related to the very weak stratification but you should mention it)

Values of ε_T were removed above 75 m, as there is an apparent large overturn in temperature at approximately 60 m depth. From the glider CTD data, it is clear that this is density-compensated and is therefore not a turbulent overturn. This will be added to the text:

"We remove all values of ε_{τ} shallower than 75 m because there is an apparent large temperature overturn above this depth that is, in fact, density-compensated; hence, this is not a real overturn."

Section 3.1. Uncertainties reported in this section have inconsistent values and units, also with respect to Figure 4. E.g. values in line 180, are much higher than those in 184 (and have no units, while those in 184 do have units), and they also seem different from the error margins in the figure. Please, clarify how the standard deviations are calculated and improve the consistency.

We thank the reviewer for pointing out these inconsistencies, in particular the incorrect use of units when discussing the dimensionless geometric standard deviation factor in line 184. We had not included a definition of the geometric standard deviation factor in the manuscript as this can be found in standard mathematical textbooks. However, given that many readers will not be familiar with geometric means and geometric standard deviation factors, we will expand the descriptions given here to provide some clarification.

Lines 131 to 134 will be altered from:

"Quality-controlled estimates of ε_{μ} were binned, profile by profile, into 25 m bins; we use the geometric mean (and standard deviation) in preference to the arithmetic mean, the better to represent the average (and spread) of observations that span many orders of magnitude."

to:

"Quality-controlled estimates of ε_{μ} were binned, profile by profile, into 25 m bins; we use the geometric mean in preference to the arithmetic mean, the better to represent the average of observations that span many orders of magnitude and which are not normally distributed."

Lines 179 to 180 will be altered from:

"The greater noise in ε_T is reflected in its having a higher standard deviation factor (5.57) than ε_{μ} (3.94). (Note that geometric standard deviation factor is multiplicative, not additive.)"

to:

"The greater noise in ε_T is reflected in its having a higher geometric standard deviation factor (GSDF) of 5.57 than ε_{μ} , the GSDF of which is 3.94. (Note that GSDF is multiplicative, not additive, and is therefore dimensionless. The one-standard-deviation-range is from the geometric mean/GSDF to geometric mean × GSDF.)"

Throughout the paper, we have clarified when we are using the arithmetic or geometric mean and standard deviation.

Line 195. There is no panel a in Figure 5.

This will be altered to simply refer to Figure 5.

Lines 195—196. Not sure I see this modest peak you talk about, or maybe it is too modest to be worth mentioning?

This sentence will be removed: we agree that the modest peak is very hard to see.

Line 199 – 200. Is there something wrong with this sentence? I do not get it.

The wrong number was used to illustrate this point. This will be corrected:

"This is in contrast to higher values of ε_{μ} and ε_{T} (> 10⁻¹⁰ W kg⁻¹) within the 700 to 800~m depth range at the beginning and end of the deployment."

Line 205. Why? (see comment above)

As mentioned above, values of ε_T were removed above 75 m, as there is an apparent large overturn in temperature at approximately 60 m depth.

Line 206—207. I think you need a bit more clarity in this sentence. What makes the epsilon and kappa distributions similar?

This will be changed from:

"The distribution of κ_{ρ} resembles that of ε_{μ} , notwithstanding a decrease in κ_{ρ} at mid-depth between the 26 and 28 January where ε_{μ} remains relatively high (Figs. 4a and 6a)."

to:

"Below approximately 250 m, the distribution of κ_{ρ} resembles that of ε_{μ} due to the relatively low variability in N² at these depths (Fig. 3c). Above 200 m N² increases substantially in the pycnocline and thus κ_{ρ} decreases."

Line 216. Epsilon units missing

Units will be added.

Line 217. Is there an upward diffusive heat flux in the upper layer? Does that mean that thermal stratification is inverse?

There is indeed a small thermal inversion near the surface – i.e. the temperature maximum (> 28° C) is found at a depth of approximately 60 m. Above this, temperature decreases by approximately 0.5° C. This is not visible in Figure 3a given the contour interval. We will explain this in the revised paper in relation to the upward heat flux visible in Figure 6b.

Line 235. "[...] and given that theoretical flux laws can overestimate κ S in the real ocean [...]". This sentence needs a reference.

References to Taylor and Veronis (1996), Kelley et al. (2003) and Radko (2005) will be added in the revised paper.

Line 245. By eye this difference looks greater. More details on how the averaging is performed are needed.

We will be changing the method by which we selected those parts of the water column in which we calculate the fluxes of heat and salt due to salt fingering (see above), so these figures and average numbers are going to change. We will provide more details on how the averages are performed – in particular including zeros in between salt fingering regions.

Line 258. "sampling and the necessary setting of a minimum Thorpe scale", why do you need a minimum Thorpe scale? (see comment above)

We apologise, the reference to setting a minimum Thorpe length scale was an error: the original text should have referred to the setting of a minimum length scale for a turbulent patch. The discussion of the influence of resolution on the reliability of ε_T estimates will be thoroughly revised, but the original error will be corrected.

Lines 262-263. "however, the lack of similarly high-resolution salinity observations means that the resolution of N used to calculate ϵT in Eqn. 5 will limit any improvement that a small minimum LT might otherwise have on estimate of ϵT ". The N that you introduce in the ϵT formula is, in my view, a background N representative of scales larger than the turbulent overturns, so I don't necessarily see the lower resolution of N as a problem here.

We agree with the reviewer that the N that we introduce in the ε_T formula is meant to be a background N representative of scales larger than the turbulent overturns. Therefore, the lack of high-resolution salinity observations is not a problem. We will remove this sentence from the revised paper.

Lines 264-265. "Howatt et al. (2021), who reduce the difference between their two estimates of ϵ ," I don't get the meaning of this sentence

This comment will be removed.

Line 285. See main comment (4)

Please see our response to main comment (4)

Lines 290-292. "The downward flux of heat due to double diffusion is almost four times that due to turbulent mixing; the downward flux of salt due to double diffusion is 7.5 times that due to turbulent mixing.". The differences in diffusivities (as reported in line 245), appear much weaker. More details are needed on this calculation.

The depth-time distributions of thermal (K_{Θ}) and haline (K_S) diffusivity do not match the depth-time distribution of turbulent diffusivity (K_{ρ}). One can imagine, therefore, a point with a large temperature gradient, Θ_z , a large K_{ρ} but a small K_{Θ}; and another point with a small Θ_z , a small K_{ρ} but a large K_{Θ}. The

ratio of average K_p . Θ_z to average K_{Θ} . Θ_z will not be the same as the ratio of K_p to K_{Θ} . We will make this point clearer in the revised paper.

Lines 298-300. "We posit that this improved agreement is due to εT being calculated using the same high-resolution observations as were used to calculate $\varepsilon \mu$: other studies have compared $\varepsilon \mu$ to εT calculated using lower-resolution temperature observations." That seems plausible but you have the data to test this and I think you should (see main comment (2))

Please see our response to main comment (2).



Figure 1. Turbulent kinetic energy dissipation rate, ϵ (W kg⁻¹), as estimated using (a) the Batchelor spectrum method (ϵ_{μ}) and (b) the Thorpe scale method (ϵ_{T}). The respective means (thick lines) and standard deviations (thin lines) are shown in the panels on the right. Note that geometric standard deviations is multiplicative. The difference between ϵ_{μ} and ϵ_{T} .



Figure 2. Exemplar spectra from which ε_{μ} was estimated. The spectrum derived from FP07 temperature observations is plotted in blue; the best-fit Batchelor spectrum is plotted in black. The polynomial fit to the observed spectra is plotted in light blue and the polynomial fit to the Batchelor spectra is plotted in red: both take the form $ax^2 + bx + c$. The parameters, as described in Section 2 of the paper, used to determine the goodness-of-fit are printed above the plot.



Figure 3. Histograms of Thorpe length scales as estimated from the FP07 fast thermistor observations (purple line) and the Seaglider's ordinary CTD observations (black line).



Figure 4. Thorpe scale, LT (m), estimated from FP07 fast thermistor observations.