## Summary

I initially refused the review request but can now provide comments. I've elected to do so as a community member rather than an anonymous reviewer. My review will be more narrated than usual, given the open peer discussion at EGU journals.

My understanding is the authors developed a "reduced algorithm" for estimating two turbulence quantities ($\varepsilon$ and $\chi$) from the voltage spectra of shear and fast-temperature sensors onboard an expendable profiler (SOLO). It needs to be clarified from the methods' description, but the $\chi$ estimates rely on first obtaining $\varepsilon$ from the shear probe. The algorithms are designed to minimize the data transfer rate by fitting a narrow range of frequencies of voltage spectra with a power relationship. This model spectra may not apply over the fitted range (as noted by the authors). Whether the observations are expected to have a power relationship over the fitted frequencies is accounted for with a correction factor $F_{Na}$ (shear) or $F_{Kr}$ (temperature-gradient). They have chosen the empirical Nasmyth (inertial subrange) and the Kraichnan models (viscous-convective subrange) instead of using an inertial model for both datasets or a viscous model for both.

Below I summarized further my understanding of the algorithm before providing more details about modifications that may render the article more transparent for readers.

### Algorithm summary

For each segment of data (e.g., $\sim$ 5 s chunk of the full profile), the algorithm estimates:

- an average and minimum drop-speed onboard the instrument
- voltage frequency spectra from each temperature and shear probe sensor (are there 2x of each onboard?)
  - Their spectra have roughly 6-10 degrees of freedom in their setup by using 3 overlapping segments of 256 samples each. The spectral bandwidth is from $\sim$0.4 Hz to 50 Hz with a frequency resolution of 0.4 Hz. The drop speed is about 0.2m/s, so their spectra cover wavenumbers ranging from 2 to 250 cpm. Given the thermal frequency response and the spatial size of the shear probe, the spectra are probably "usable" up to 20 Hz (thermistor) and 100 cpm for the shear probe. Of course, noise can limit this further. Still, the drop speed of 0.2m/s nicely "optimizes" the usable range in both the shear and temperature gradient spectra.
  - The calibration coefficients are not stored on the SOLO, unlike recoverable turbulence profilers or most ocean sensors. The lack of calibration constants prevents them from converting the voltage spectra into physical units onboard the SOLO (or onshore since no spectral observations are transmitted).
- Two power fits are performed for each voltage spectrum over a narrow bandwidth of available spectral observations. This bandwidth is between 1 and 5 Hz (1/2 decade of data). The first power fit is between 1 and 3 Hz, and the second is from 3 to 5 Hz. Overall, their 1/2 decade has 10 spectral samples, and each power fit is done with 5 samples.
- Two quantities, one for each power fit, is returned to shore for each voltage spectrum. These quantities are then converted into initial $\varepsilon_{\text{init}}$ and $\chi_{\text{init}}$ estimates via the calibration constants of the sensors.
- From these initial estimates, the correction factors $F_{Na}$ and $F_{Kr}$ are used to obtain $\varepsilon$ and $\chi$. I presume that $\varepsilon$ is also fed into $F_{Kr}$ to obtain $\chi$ (not clear from the methods' description). Both correction factors depend on the frequencies fitted, the choice of model, etc. It's unclear whether $F_{Na}$ and $F_{Kr}$

are sufficiently general such that any frequencies could be used for fitting the voltage spectra (e.g., different profiling speed or source of vibrations).

- There are accelerometers onboard the SOLO, but they are not used to correct the turbulence (voltage) spectra. Instead, the analysis utilizes data between 1 to 5Hz to avoid surface waves and motion-contamination. The authors state the accelerometers are for computing wave statistics, which I presume will involve additional spectral computations onboard the SOLO. How the turbulence analysis changes in wavy flows dominated by surface waves should be discussed in the ms.

## Major Comments

I have grouped my concerns into three themes. The main suggestions for implementation in the ms are numbered and shown in *italic purple*.

1. More transparency is required in discussing the drawbacks of their chosen strategy. The algorithm was designed to limit data transmission at all costs by relying on a small subset of data recorded by the profiler.
2. Lack of assessment of the fit-score as an alternative measure of data quality.
3. Lack of transparency about the $\chi$ methods and the data quality both in applying the reduced algorithm and for estimating $\chi$ from standard practices.

# 1 Reduced algorithm's "framework"

## 1.1 Improvement over using band-averaged (low-resolution) spectra

The ms should be more transparent in explaining the drawbacks of their chosen strategy when compared to that employed by *Rainville et al.* (2017). As noted by the authors on L27-31, the reduced algorithms of *Rainville et al.* (2017) sends band-averaged (i.e., low-resolution) spectra ashore. In their case, 12 spectral observations are transmitted that cover all available lengths and timescales of turbulence, including the noise (a bit overkill, in my opinion). Specifically, their instrument sends 9 spectral samples between 2 to 100 cpm and 4 samples over the wavenumber range (5 to 25cpm) used in the current ms for fitting The main advantage of sending spectral observations ashore is being free to apply standard fitting or integrating techniques to estimate the turbulence quantities of interest. More importantly, quality-control criteria, such as the mean absolute deviation listed by the authors on L240 can be calculated. This criterion indicates whether the observations follow the expected forms of turbulence, i.e., spectra aren't drowned by noise, motion contamination, or anisotropy. These issues are usually also nicely spotted by inspecting spectra, but with the proposed reduced algorithm in the ms, this information is lost during transmission.

*1. Since the authors only used limited range of turbulence scales (1 to 5Hz), it would be worth highlighting the data "savings" that they gain by transmitting two power fit estimates (1-3Hz, and 3 to 5Hz) over sending for example 4x (band-averaged) spectral observations.*

*2. It would also be useful to the say why a data reduction scheme that reduces 512 samples to 12 samples (band-averaged spectra) is inadequate, which ultimately resulted in developing an algorithm that fits a narrow range of information onboard the processor.*

### 1.2 Using a limited range of frequencies for fitting

The algorithm appears highly dependent and applicable for their particular drop speed, sampling rates and frequencies used for fitting. It also depends on there being no motion contamination over the range of frequencies used. It makes the paper highly specific for their platform, as opposed to being a "reduced algorithm" for turbulence profilers. This is fine but worth highlighting. However, the reliance on a small subset of the available turbulence information is problematic. Makes you question why even sample the shear signal at rates above 32-64Hz if we can get away with deriving $\varepsilon$ (or $\chi$) from such a narrow range of turbulence length/time scales.

Furthermore, the narrow band of scales used comprises a low number of spectral observations (5x per fit), which have low statistical significance given how the voltage spectra are calculated. The spectra use 3x FFTs with a 50% overlap, each having 256 samples, which is less than typical when accelerometers are used to decontaminate the spectra. Depending on the widowing function applied, this is about 6 to 10 degrees of freedom – a tad more than spectra with no statistical significance. The degrees of freedom wouldn't allow using squared-coherency to decontaminate the spectra (a minimum of 7 NFFTs would be more appropriate). There's all this effort to reduce the data transfer, but an easy saving would be using a statistically significant spectral averaging strategy, e.g., 768 samples for each segment instead of 512 samples (the 5x NFFTs would still have the exact resolution).

*3. Would the results be similar if a more statistical significant spectra were used in the calculation? Would band-averaging the spectra change the estimated $\varepsilon_{init}$ ($\chi_{init}$)?*

*4. Fig 3: Include the kinematic viscosity and confidence intervals for these spectra (see §5.4.8* Emery and Thomson*, 2001) for calculating the confidence levels*

### 1.3 Number of samples used for each fit

Recent work has concluded that for data with low variance, 8 samples are required for regressions (*Jenkins and Quintana-Ascencio*, 2020). For highly variable datasets, this number increases to 25. The authors have used 5x spectral observations, compounding the above issue of relying on low bandwidth of turbulence spectral observations that each have low statistical significance. It makes you wonder if the least-square power fit is just a guesstimate of what the turbulence level in the signal might be.

*5. How large are the confidence intervals for the fitted quantity in Eq 23? Least-square fitting usually allows for this result to be calculated, but these have not been presented.*

*6. What would be the confidence level for $\varepsilon_{init}$ if the spectral confidence levels were propagated, along with the errors associated with using only 5x samples for a least-square fit? The error would propagate through to $F_{NA}$ in Fig 2.*

### 1.4 No reduction of vibration and wave-contamination

The ms mentioned using accelerometers to determine the wave climate, so spectra are presumably being calculated. It seems no information gained from the accelerometers will be used to assess the quality of the turbulence measurements? Rather than decontaminate the shear spectra, the fit is restricted to frequencies 1 to 5Hz when some platforms do have contamination across those ranges (Fig 5 of *Bluteau et al.*, 2016). Surface waves "seas" are awfully close to the frequency range used here.

*7. The authors should acknowledge that their chosen frequencies avoids motion-contamination for their specific platform.*

*8. It would help to detail in the ms why it's unfeasible to perform this calculation onboard the processor. A squared-coherency estimate (*Zhang and Moum*, 2010) is mostly rearranging cross-spectral terms (conjugates of FFTs). It is not restricted to being done in physical units. This additional processing could indeed be accommodated by using longer segments. We often live with ADCP profiles with 10 m bins, so getting a lower vertical resolution turbulence signal might be worthwhile if the estimates are more robust.*

## 2 Fit-score as a quality-control indicator

The authors' "fit-score" is the ratio of the result obtained using the first 5 samples (1 to 3 Hz) vs the subsequent 5 samples (3 to 5Hz). This quantity tests crudely the sensitivity of the results to a very slight change in frequencies. Having used it myself as a qualitative guide (see Fig 4 of *Bluteau et al.*, 2011), I'm concerned that there's no measure of whether the entire spectra are “garbage” particularly the temperature gradient spectra that are usually much more variable in quality.

If the turbulence shape varies widely over such small changes in frequencies, then, indeed, the data is very poor. The question, though, becomes, how poor? How much variations in shape can we expect? Does this variation depend on the spectral averaging, i.e., the statistical significance of the spectra? Will the fit-score depend on the number of samples used in the individual fits? How do all of these factors translate into rejection criteria? It needs to be clarified if a rejection threshold was proposed. What's evident in the manuscript is that the fit-score it's now the primary way to assess quality, given that the reduced algorithm does not send spectra observations to shore.

*9. To develop and assess the algorithm, the SOLO was recovered. Thus, it's possible to estimate the mean absolute deviation listed on L240. The mad has existed for 22 years and is on its way to being recommended by the SCOR working group #160 as a quality-control indicator for data archiving. The fit-score should be compared to the mean absolute deviation (mad) for all segments in a scatter plot. The scatter plot would enable readers to judge the robustness and usefulness of the fit-score as a quality-control indicator. Let the data speak for itself.*

## 3 Data quality of the temperature gradient data

### 3.1 Methods description

In general, the $\chi$ description is unclear. The ms is organized as if the $\chi$ estimates are done in isolation of $\varepsilon$, when Eq 30-31 shows that $\varepsilon$ is required to estimate $\chi$. This strategy differs from the many previous chipod papers (e.g., *Moum and Nash*, 2009; *Becherer and Moum*, 2017), which use the fast-temperature data to solve for both $\varepsilon$ and $\chi$ by equating Osborn to Osborn-Cox's model. The proposed algorithm design in the ms was justified by referring to chipod papers pioneered by their research group (*Becherer and Moum*, 2017). However, the authors haven't highlighted that the new algorithm depends on getting $\varepsilon$ first from the shear probe before obtaining $\chi$ from the temperature gradient spectra rather than obtaining $\varepsilon$ and $\chi$ simultaneously from the temperature gradient spectra. Needing both shear probes and fast-temperature sensors in itself increases the amount of data processing and data transmission from the SOLO. Why not get rid of the shear probes completely? Does the SOLO not measure background temperature and salinity?

*10. Please explain in the intro [L35-40] whether they are using the shear probe to derive $\varepsilon$, and if this quantity is then used for estimating $\chi$. Some comments as to why this strategy was chosen should be provided given it increases the demands on data processing and data transfer.*

*11. The ms could also better discuss the implications of relying on $\varepsilon$ from the reduced-algorithm on the quality of the $\chi$ estimates. Unless of course, they've estimated $\varepsilon$ and $\chi$ simultaneously from the temperature gradient spectra in which case the ms should illustrate how $\varepsilon_\chi$ compares to $\varepsilon$ obtained from the shear probes.*

*12. Also re-iterate the dependency of $\chi$ on $\varepsilon$ when presenting Eq 30-31, and on L306 when claiming the temperature gradient spectra $\Phi_{Kr}$ depends only on $\chi$, which isn't true (see Fig r1).*

### 3.2 Information about the chi data quality

The fact that no temperature gradient spectra were shown in the ms is disconcerting. A few shear probe examples focus on observations between 1 and 5 Hz, which mask any spectral contributions from waves or motion contamination. The only spectral information for $\chi$ is the data density plots in Fig 10, which only show observations between 1 and 5Hz. Still, this figure gives an inkling of the temperature gradient data quality. Fig 10b shows a cloud of data with the wrong slope sign at non-dimensional $k \lesssim 3 \times 10^{-2}$, which then starts to fall off too early. Is this because of problems with the estimated $\varepsilon$ shifting the spectra to the left? Fig 10c looks like spectra drowned by noise (perhaps high $\varepsilon$ and low $\chi$). Even the high-score examples in Fig 10s could be better. The "peak" data density doesn't fit the Kraichnan form. There's no curvature in the location of maximum data density, unlike Fig 7a for the shear probe.

I'm questioning whether it's the algorithm behaving poorly or whether the collected temperature gradient data could be of better quality at the outset.

*13. To alleviate concerns about the data quality, I strongly recommend adding an extra column in Fig 7 and 11 showing the data density for all wavenumbers, not just those used by the algorithm.*

*14. Another request is adding an extra column in Fig 4 with the temperature gradient spectra collected concurrently with the shear probe. Preferably all available wavenumbers for data transparency. There are none in the ms, which masks the data quality.*

*15. Specify which $\chi$ estimates were rejected from the paper. Fig 10 shows variable fit-score, but which would be flagged as unusable for further analysis in a scientific article?*

*16. Why use the inertial subrange model for shear but then assume the viscous-convective subrange model for temperature when the same frequencies are fitted onboard the processor? Does it not matter that the fitted model isn't expected over the range k.*

*17. Also, please use a colour scale/scheme that compares the temperature and shear spectra data density against each other. The colour gradient presents a data count (or proportion of data), but the scheme changes between figures. If a colour theme for shear and temperature is necessary, decorate the x and y-axis colours, but leave the colour gradient the same across all the data density plots (Fig 6, 7 and 8-9).*

*18. The colours in Fig 6, 7 and 8-9 also seem to saturate at values below the maximum, which makes it hard to see where the maximum sits relative to the theoretical shape (Fig 7-10) and the 1:1 slope (Fig 6 and 9). Perhaps visit brewermap or cmocean color palettes described in* Thyng et al. *(2016) article about ocean data visualization.*

## OTHER COMMENTS

### 3.3 Misciting

Remove the erroneous citations to my published work. On L342, the reference to my article on fitting shear probe data does not state that centring a fit around 10-20 cpm minimizes sensitivity to the fitted range. The only thing that reduces this sensitivity is fitting the correct model (e.g., inertial subrange) over the wavenumbers that this model is expected to be valid. Another way to limit sensitivity is to have high-quality measurements that aren't drowned by noise, vibrations and surface waves. The range of wavenumbers would depend on data quality, drop speed, and model used for fitting. If my results were insensitive to the 10-20 cpm range, I was using a model that covered both the inertial and viscous subranges, and the data was of good quality after decontamination.

L344-349. I'd remove the whole paragraph. First, the power fit used by the authors sometimes uses frequencies in the inertial subrange (1 to 5Hz translates to 5 to 25cpm). With the 0.2m/s drop speed, the inertial subrange is only being fitted with the correct model when $\varepsilon > 10^{-7}$ W kg$^{-1}$. For low $\varepsilon < 10^{-7}$ W/kg, the inertial subrange model is fitted to the shear probe's viscous subrange. A similar argument would apply for temperature, except that the authors have chosen to always apply the viscous-convective model instead of the inertial-convective model (Fig. r1).

Using a moored platform changes very little other than we have to contend with variable speeds past the sensor. A mooring doesn't automatically translate into long FFT segments. My miscited article (*Bluteau et al.*, 2011) refers to the spectral fitting of the inertial subrange of acoustic-velocity measurements – not shear probes. We use long segments because we're relying on the lower scales of turbulence since the data quality is typically too poor over the viscous subranges (technological issue). Another reason for using long segments is calculating other turbulence quantities such as Re stresses and TKE. These estimates rely on covariances and thus the integration of cospectra with reasonable statistical significance (need more NFFTs, and/or band-averaging).

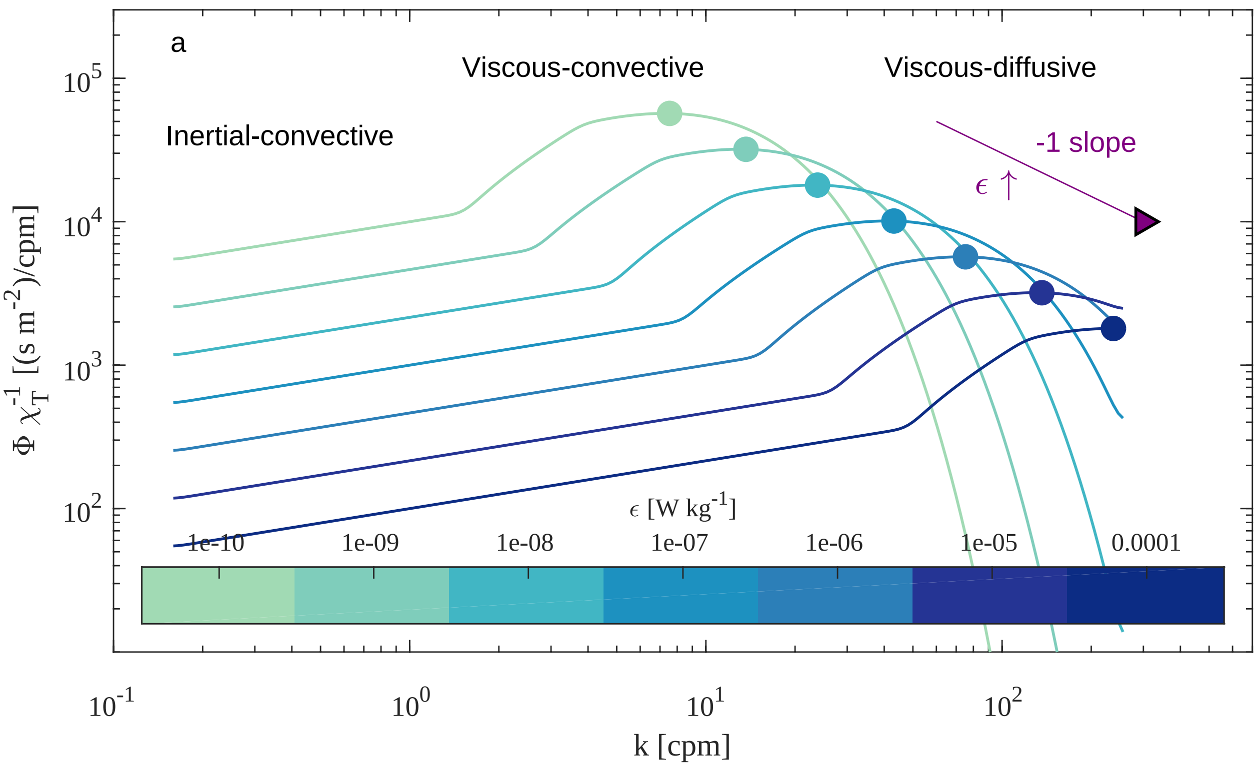


**Fig r1:** Example of how the temperature gradient spectra changes with $\varepsilon$ (along the -1 slope) and with $\chi$ (vertically). The current ms uses turbulence information for wavenumbers $k$ from 5 to 25 cpm (frequencies 1 to 5 Hz).

## References

Becherer, J., and J. N. Moum (2017), An efficient scheme for onboard reduction of moored chipod data, *J. Atmos. Oceanic Technol.*, *34*(11), 2533 – 2546, doi:10.1175/JTECH-D-17-0118.1.

Bluteau, C. E., N. L. Jones, and G. N. Ivey (2011), Estimating turbulent kinetic energy dissipation using the inertial subrange method in environmental flows, *Limnol. and Oceanogr.: Methods*, *9*, 302–321, doi:10:4319/lom.2011.9.302.

Bluteau, C. E., N. L. Jones, and G. N. Ivey (2016), Acquiring long-term turbulence measurements from moored platforms impacted by motion, *J. Atmos. Oceanic Technol.*, *33*(11), 2535–2551, doi:10.1175/JTECH-D-16-0041.1.

Emery, W. J., and R. E. Thomson (2001), *Data Analysis Methods in Physical Oceanography*, 2nd ed., 638 pp., Elsevier Science, Netherlands.

Jenkins, D. G., and P. F. Quintana-Ascencio (2020), A solution to minimum sample size for regressions, *PLoS One*, *15*(2), 1–15, doi:10.1371/journal.pone.0229345.

Moum, J. N., and J. D. Nash (2009), Mixing measurements on an equatorial ocean mooring, *J. Atmos. Oceanic Technol.*, *26*(2), 317–336, doi:10.1175/2008JTECHO617.1.

Rainville, L., J. I. Gobat, C. M. Lee, and G. B. Shilling (2017), Multi-month dissipation estimates using microstructure from autonomous underwater gliders, *Oceanography*, *30*, 49–50, doi:10.5670/oceanog.2017.219.

Thyng, K. M., C. A. Greene, R. D. H. ond Heather M. Zimmerle, and S. F. DiMarco (2016), True colors of oceanography: Guidelines for effective and accurate colormap selection, *Oceanography*, *29*, 9–13, doi:10.5670/oceanog.2016.66.

Zhang, Y., and J. N. Moum (2010), Inertial-convective subrange estimates of thermal variance dissipation rate from moored temperature measurements, *J. Atmos. Oceanic Technol.*, *27*(11), 1950–1959, doi:10.1175/2010JTECHO746.1.