Response to Referee 2

We appreciate the time taken by the referees to review our manuscript and provide comments and suggestions for its improvement. We have taken some time to essentially reset and rewrite the manuscript which, although the measurements and results are largely unchanged from our initial submission, has resulted in what we believe to be a much cleaner submission with more direct focus on the measurement technique and much less attention spent on interpretation of the results.

Below, we respond to the individual comments made by the referee on the last version of the manuscript. To do so, we have reproduced the original review, with our comments provided in blue text.

The authors' reply clarified points that helped me to better understand what had been done. The amount of work involved is indisputable, but some questions remain as to the reliability and interpretation of some results. Also, at this level of evaluation, more attention to form would have been appreciated, because there are many avoidable presentation errors. This gives the impression that the manuscript has been hastily revised and not proof-read. For example, there are many typos in the added text, and, e.g., in equation (2). Figure 9 caption does not correspond to the new version of the figure. The manuscript has been improved but, from my point of view, still requires major revisions before publication.

The referee is correct, the previous revisions of this manuscript were hurried, leading to numerous mistakes and inconsistencies that we have done our best to fix in this version.

Comments:

– The wind profiles measured from the onboard GPS during the ascent (now shown in figure 8) are very important added values and provide convincing arguments for the quality of wind profiles measured by Hydron. They provide much more important information than NWS radiosondes (low resolution and large distance). However, they are only used for wind profile comparisons in Figure 8 and the comparison results are not mentioned in the abstract and conclusions. These additional GPS profiles should provide a good reference and can also be used to estimate wind shear at a vertical resolution of ∼100 and ∼1500 or 2500 m, since they are not affected by the (large) horizontal excursion as the Hydron. The wind comparison result is particularly good above 10 km for flight 3, with wind fluctuations of similar vertical wavelengths and amplitudes. Contrary to the authors' claims, this seems to suggest that the large horizontal excursion of the Hydron (and horizontal inhomogeneity) may not be the dominant factor explaining the shear (and likely N2) fluctuations at the vertical sampling of 100 m from time series. Also, the shear profiles in Figure 9 obtained at the vertical resolution of 100 m and supposed to be strongly "influenced" by the horizontal excursion (1500 m) are quite typical of shear profiles estimated from balloon-borne radiosondes at a vertical resolution of 100 m (after applying a low-pass filter with a cut-off at 200 m on 10-m vertical resolution profiles). See the example below from a Vaisala radiosonde. This is another reason why I remain skeptical about the interpretation of the results (lines 448-449) shown in figures 9 and 10, in terms of horizontal inhomogeneity.

  Although we still believe that there is some horizontal heterogeneity picked up in these measurements, for the sake of expediency the revised manuscript no longer addresses horizontal heterogeneity (except as a potential complication when interpreting the results).

  In addition, if the effective vertical resolutions of Ri_t and Ri_z (and shear) are 100 m and 2500 m, respectively, the large difference between the two (figure 10) can be due to the (vertical) scale dependence of Ri (and shear) (See: Balsley, B. B., Svensson, G., and Tjernström, M.: On the scale dependence of the gradient Richardson number in the residual layer, Bound.-Lay. Meteorol., 127, 57–72, 2008.). Therefore, the difference in fluctuations at the two resolutions cannot be attributed to the (sole) horizontal inhomogeneity.

  It turns out that I do not agree with the statement lines 481-485 because the authors refer to the estimates of Ri_z, always »1 due to the poor vertical resolution (∼2500 m). It is not adapted for the detection of shear instabilities commonly observed in the tropo-stratosphere (∼a few hundred meter deep or less).

The fact that the vertical sampling is the same for the two profiles (Figure 9-10) is misleading: it gives the impression that the resolution is the same. The effective resolution should be given by using a clear terminology. For example: R_{100} and Ri_{2500}. It should be used everywhere. Line 648 ("... with high Ri...") is ambiguous because the calculated Ri's cannot be directly compared to the thresholds (0.25 or 1).

We agree that the large difference in the magnitude of Ri estimates are due to the large differences in vertical resolution. We note that the aircraft measurements will be much more sensitive to the horizontal gradients than to the vertical gradients due to the shallow glide slope, which could be contributing to the fluctuations in S^2. However, in the revised version we have trimmed down the section on Richardson number calculation and no longer discuss variability in the results as relating to horizontal heterogeneity.

– Lines 410-415: It is not clear what the authors mean. The larger vertical wavelength of the fluctuations should be favorable to larger horizontal scale, and thus to horizontal homogeneity. As a result, the aircraft's orbital trajectory should be less problematic under such conditions. In addition, the argument is weak: the wind profile during flight 2 shows larger amplitude fluctuations during the balloon measurements (ascent) between 5 and 8 km than during the Hydron measurements (descent).

'the larger vertical wavelength of the fluctuations should be favorable to larger horizontal scale, and thus to horizontal homogeneity' would require isotropy in the velocity fluctuations, which is not likely at these large scales.

– In line 409, do the authors mean "...that are NOT evident in the NWS soundings..."?

This was indeed a typo that should have read 'than are evident in the NWS soundings' instead of 'that are evident in the NWS soundings'.

N is shown instead of N^2 in figure 9. The profiles contain negative values, which is not possible, because N is necessarily real positive (when $N^2 > 0$) or imaginary (when $N^2 < 0$). Can the authors explain how they calculated N in practice?

We failed to mention in the previous version that we had preserved the sign of N^2 when presenting N, i.e. $N = \mathrm{sgn}(N^2)\sqrt{|N^2|}$. To avoid this ambiguity in the current revision we have reverted to presenting N^2 and S^2 instead of N and S.

– Lines 483-484 are not clear.

The revised manuscript no longer includes this statement

– Line 505: "...caused by inertial turbulence...".

This depends on your definition of turbulence. For example, laminar flows can certainly provide non-zero values of k as typically defined (e.g. laminar vortex shedding, laminar waves, low speed velocity fluctuations introduced by changes in boundary conditions, etc. will all produce oscillations/unsteadiness about the mean value that can produce $k > 0$). If one defines *all* unsteadiness as being turbulence, then the suggested revision would be correct. However, this is not true in the strictest sense and the original statement that "elevated values of k may not necessarily correspond to velocity fluctuations caused by turbulence." is more general and therefore we elected to keep it in place.

– Line 542: n must be equal to -5/3.

Only if we are assuming all locations contain inertial turbulence. As was stated later, the fitted value of n being close to -5/3 ($\pm 10\%$) was used as a metric to identify the presence of inertial turbulence. However, this is not relevant as the revised manuscript no longer attempts to distinguish between turbulent and non-turbulent events and forces the fit to use $n = -5/3$ to simplify the calculation of ε.

– In Figure 12, the red lines should be limited to $f > 0.1$ because it was the threshold used for linear fitting.

The revised manuscript no longer includes these figures.

– Line 595: "... nearly constant with α" (?)

The revised manuscript no longer includes the figure being referred to.

– In the conclusion section, Lines 646-649 comment on comparisons between Ri and EDR (the former figure 13), which are no longer described. Qualitative comparisons can only be made from Figures 14-15 and 16.

The revised manuscript has a completely rewritten Summary and Conclusions section.

Response to Referee 3

We appreciate the time taken by the referees to review our manuscript and provide comments and suggestions for its improvement. We have taken some time to essentially reset and rewrite the manuscript which, although the measurements and results are largely unchanged from our first submission, has resulted in what we believe to be a much cleaner submission with more direct focus on the measurement technique and much less attention spent on interpretation of the results.

Below, we respond to the individual comments made by the referee on the last version of the manuscript. As the referee's comments to the last revision was appended to our response of their review of our original submission, we have only included comments referring to the most recent submission. Furthermore, we have preserved the entire comment/response chain below. To try to minimize any confusion regarding the sequence of the discussion, we have included our initial response in green and our current responses are included in blue. We hope that this is not too confusing.

1. Similarly: "which allowed the connection to be made between the locations of increased turbulence intensity and the source of its generation" was tenuous, only to the level of "consistent with".

   This is a fair assessment, we have revised the text accordingly.

   The abstract still contains this same phrase.

   We have rewritten the abstract in the revised manuscript.

   Also in the abstract, the sentence "By being able to transect the air, the glider allows for turbulence wavelengths to be sampled at a particular altitude, improving statistical convergence and spatial resolution of derived statistics from its in-situ sensors." is misleading, since the glider cannot remain at a particular altitude, and demonstration of improvement of convergence and spatial resolution remains missing in the paper, as noted in the previous comments above.

   We have rewritten the abstract in the revised manuscript. A more nuanced response to this comment is also detailed below.

2. "However, due to the transient nature of their Lagrangian flight trajectory, balloon-based approaches are not necessarily amenable to obtaining detailed statistical descriptions of turbulence at high altitudes." Why? Aircraft are also transient, and if GPS guided, only see what is advected past. Balloons with altitude profiling are not Lagrangian vertically, so also sample more than one parcel of air. The statement is vague: it depends what statistics are being evaluated.

   True, and addresses the same point as the first detailed comment. We have revised the statement to be less vague.

   115: "However, as balloons advect horizontally with the wind and are unable to maintain a fixed geospatial location, balloon-based approaches would benefit from complementary measurement approaches" is even more vague.

   See below regarding Eulerian vs Lagrangian observations.

   120: The scientific advantage of remaining over a fixed ground location is not self-evident, as the features of the atmosphere typically do not. Or is the advantage primarily operational in nature? Please articulate this point. Also, the transecting ability of a glider is not related to the helical profile, but to its horizontal airspeed. The sentence "These qualities facilitate the statistical analysis necessary for quantification of non-stationary properties." remains unsupported in the paper (that I can find).

   We were not aware that this was not self-evident. Particularly since most (if not all?) numerical models of the atmosphere are conducted on Eulerian grids, it would seem intuitive that observations intended to improve our ability to model the atmosphere should therfore be conducted as close to a fixed geospatial reference location as possible (i.e. the UAS stays within a spatial scale corresponding to a numerical grid dimension). Furthermore, studies of turbulence in a Lagrangian frame of reference introduce different statistical behavior than Eulerian studies of turbulence (see, for example, Federico Toschi and Eberhard Bodenschatz (2009) "Lagrangian Properties of Particles in Turbulence" Annual Review of Fluid Mechanics 2009 41:1, 375-404). For example, it has been long recognized that Eulerian and Lagrangian integral scales are not interchangeable when modeling turbulent dispersion.

90: “A glider offers advantages over traditional balloon launches by being able to maximize time at altitude during its descent phase” Vertical rates for the glider vary from 5 m/s to 1 m/s, very similar to descending balloons. “These qualities facilitate the statistical analysis necessary for quantification of non-stationary properties” is not supported by evidence in the paper.

We respectfully, but strongly, disagree on this point. Note that in 1000 m of altitude change, for the current experiments the orbit of the glider means that it samples approximately 15,000 m along its flight path, whereas a balloon will sample only the 1000 m. This is a significant difference in the amount of atmosphere and range of eddy sizes that are sampled over the same vertical distance. Note also that the current configuration of the aircraft means that the turbulent eddies are acquired at an order of magnitude higher temporal resolution as well. It therefore would be very difficult to reproduce the spectra of Figure 11 (and corresponding k and ε estimates), and azimuthal distributions of these statistical properties (as shown in Fig. 14) with a balloon, particularly over the wavenumber range and at the vertical resolution that the glider can measure. We have added the above discussion to the conclusions to ensure that these points are not overlooked by other readers.

The objectionable phrase has apparently been removed from the introduction. The confusion arose because the vertical resolutions of both platforms are similar; the differences relate to the horizontal velocity.

Some confusion remains in conflating stationarity with homogeneity. Variations seen in a fast (e.g., 60 m/s) transect are likely to be due more to spatial inhomogeneity than temporal changes in statistics, provided that the overall time interval of the record is small. The higher glider airspeed speed allows larger spatial scales to be assessed in a shorter time, making the Taylor assumptions used in the turbulence estimation frequency-to-wavenumber conversion more likely to hold for the larger scale observations.

We agree. But we also note that most fundamental theories of turbulence are constructed in the wavenumber/spatial domain, hence there are intrinsic benefits to measuring turbulence in the spatial domain compared to the time domain (particularly as there is reduced reliance on Taylor’s frozen flow hypothesis). True, this does require an assumption of horizontal statistical homogeneity in the vertical distance traversed by the aircraft while gliding, but this assumption is no less restrictive than the assumption of stationarity of statistics over the same vertical distance when examining behavior in the time domain. Furthermore, not only are larger spatial scales assessed in a shorter time, but also more of the smaller spatial scales are assessed. As noted, a balloon ascends/descends at roughly the same speeds as the glider, hence this is the rationale behind our statement that there will be increased statistical convergence (better averages of the scales smaller than the horizontal distance traversed.

Because the glider transects are mostly horizontal, the changes measured are likely to be due more to horizontal inhomogeneity than vertical inhomogeneity. Balloons provide vertical profiles of parcels and cannot sample laterally. Their ability to ascend/descend slowly (few m/s) provides high spatial resolution (vertically). A glider with 60 m/s airspeed and a 15:1 glideslope, although descending at a similar rate (4 m/s here) could be considered as having a similar vertical resolution, provided that the variations observed are dominated by vertical variations, and not the lateral ones. E.g., for the 1500m record lengths used, is it more likely that data reflects the 100m vertical change, or the 1500m lateral change? The latter is supported by later discussion in the paper (around line 449), where time successive vertical gradients are erratic, due to the large lateral changes, and where true vertical gradients are assessed by considering segments on successive turns on the helix spaced (vertically) about 2.5 km apart, leading to vertical resolutions on this order. This does not compare favorably to balloon vertical resolution.

A glider provides complementary measurements by sampling (mostly) laterally, and in this case with a very large helix diameter (10km) with an ability to evaluate large scale lateral inhomogeneities.

We agree on these points as well. The only thing that we should have clarified is that our comparison assumes horizontal homogeneity of averaged statistics. We are not fully convinced that this assumption is valid, however we have had some difficulty convincing other referees that some of the features we are seeing in the data can be attributed to horizontal heterogeneity at large scales. This is something that we hope to investigate further in follow-on investigations.

Here with record lengths of 1500m, lateral spatial resolution, e.g. of turbulence quantities, is of this order.

In our experience, resolution typically refers to the smallest scales that can be resolved, and not the largest scales that can be resolved. This difference in interpretation may explain some of the confusion with regards the statements in the previous manuscript version.

Higher temporal resolution due to higher airspeed is not an inherent advantage, since the frequency-to-wavenumber conversion makes, say, epsilon, invariant. The effect on instrumentation, is, in fact unfavorable, since higher sample rates are required, and the inherent bandwidth of the sensors and plumbing can be a limiting factor for turbulence parameterization.

Note that it has long been understood that Taylor's frozen flow hypothesis is a poor assumption for large scales (Moin P. Revisiting Taylor's hypothesis. Journal of Fluid Mechanics. 2009) hence the frequency-to-wavenumber invariance may not be strictly true and measurements where it can be avoided/minimized should advantageous. We do agree that higher relative velocities requires higher bandwidth sensors, hence our efforts to quantify the bandwidth of the sensor used here. We would like to point out, though, that many in-situ velocity sensing devices have a nonlinear response to velocity and hence it is actually advantageous for sensitivity and signal-to-noise ratio for relative velocities to be higher rather than lower. This is the case for five-hole-probes such as used in this study, for which $\Delta P = 0.5\rho\Delta V^2$. Hence, a ± 1 m/s velocity fluctuation results in a 144 Pa pressure difference at 60 m/s relative velocity whereas a ± 1 m/s fluctuation at 2 m/s relative velocity results in only a 5 Pa pressure difference being produced. Constant temperature anemometers also have a nonlinear response but tend to be more sensitive at lower velocities. However, they have other issues including increased sensivity to electrical noise, calibration drift, and require calibration typically using a pitot probe (at least in wind tunnel studies).

E.g., with a 20 Hz sensor bandwidth, scale sizes greater than 2m can be observed at 60 m/s (assuming no noise floor issues). Preserving 1 decade of inertial range would require a record length of at least 20m, but with the noise floor seen in Figure 12 c), about 60m records would be required, providing a maximum of 60m (mostly horizontal) resolution for turbulence parameters. A balloon moving at 4 m/s (with the same sensors) would have a (vertical) spatial resolution of 4m.

We are confused a bit here, but believe it caused by differences in how we define resolution. As noted above, the bandwidth disadvantage at higher velocities can be offset by higher signal-to-noise ratio and lower relative velocities can actually be disadvantageous depending on the type of sensor used.

Again, from our perspective, being able to sample more wavelengths over the same 4 m vertical ascent distance is an advantage, not a disadvantage, for reasons noted above. For a balloon to capture the same wavelength range over that 4 m of ascent would require the same relative velocity as the glider, i.e. 60 m/s relative winds assuming the validity of the frequency-to-wavenumber invariance assumption. For 10 m/s relative winds, it would take 6 times longer to sample the same wavelength range (i.e. a decade) of turbulence. For the same rise/descent rate, the balloon would therefore have travelled 24 m, compared to the aircraft's 4 m, thus resulting in higher resolution in the aircraft measurement. Of course, again, this assumes horizontal homogenity in the turbulence statistics. This comparison of course assumes that a balloon is rising/descending and not neutrally buoyant, for which very long wavelengths can be captured at a single altitude, something a glider UAS could not do.

Thus, high-speed slant-path sampling has complex trade-offs that are not served by oversimplification. The paper could be improved by a more concrete discussion of the relative merits in fundamental terms, or refrain from vague comparisons.

Although we are enjoying the discussion here, our initial submission assumed that many of the points we made above would be intuitive for the readership of AMT and therefore not require lengthy exposition. Detailed evidence-based demonstrations of these points would require different experiments than the ones conducted here, which motivates future efforts. Therefore, for the sake of expediency (as this manuscript has languished overly long in review), we have elected to take the second option with the current revision and removed any vague comparisons to balloon-based measurements.

E.g., I don't see where the analysis records are defined for the Turbulence analysis. So it's hard to place the unquantified characterizations such as "high vertical resolution" (now in the Conclusion) in context. Likewise, "increased statistical convergence" would depend on a larger number of points in the data record, together with an assumption of statistical stationarity. The data record sizes were only discussed as covering a fixed horizontal distance, so the record sizes would be smaller at the higher altitudes, and the time intervals shorter. It would be helpful to describe these details in order to support the advantages discussed.

These points are largely addressed above. Specifically, we consider 'high vertical resolution' as being the large wavelength range that is sampled for a comparatively small vertical distance. We also consider 'statistical convergence' from the perspective of the number of samples of a particular wavelength of turbulence that can be sampled over a particular vertical distance (e.g. a 60 m sample would be able to average more 1 m wavelengths than a 10 m sample). The UAS also provides advantages for 'statistical stationarity' as it can sample more wavelengths in a shorter period of time, therefore reducing the time the turbulence has to evolve temporally. However, we acknowledge that if we were to define these terms with respect to the sheer number of samples acquired in time, rather than wavelengths, the data record size is conflated with the relative velocity and such comparisons become more complex.

3. Difficulty of conducting UAS measurements of this type in the NAS was not discussed, nor the conditions under which the reported flights were allowed. Was this in restricted airspace? Under who's auspices? Or was this in the NAS under a COA?

The reviewer raises a good point as the current regulatory environment prevents these types of measurements from being routinely conducted. In the current experiment, flights were conducted in restricted airspace managed by the SpacePort America facility and coordinated with the nearby White Sands Missile range. We have added revisions to the manuscript to include these points.

I don't see where White Sands coordination is mentioned.

We did not include it as we felt it was superfluous to the fact the flight was conducted in restricted airspace used. However, we have added this note into the revised manuscript.

4. how was aircraft sideslip angle determined? How did the use of this affect the quality of the horizontal wind measurements?

While revisiting the wind measurement procedures, it was found that the probe was actually rotated 90 degrees relative to what the authors initially thought. This meant that it was actually the pitch holes that were disconnected for flights 2 and 3 and not that yaw holes. We also compared flight 1 data with and without the revisions required to calculate winds for flight 2 and found that the differences were negligible. The text has been revised the text accordingly.

The discussion of wind calculation for flights 2 and 3 in paragraph 254-262 is vague and disconcerting. The pressure lines disconnected are not specified in the text. The use of vertical aircraft speed from the "variometer" (presumably pressure altitude rate, i.e., inertial aircraft vertical velocity) does not reliably indicate vertical airspeed, just as inertial horizontal velocity does not indicate horizontal airspeed. "Flight 1 data was processed with the original and revised approach, and the impact on the results on the resulting wind velocity statistics found to be negligible". Which statistics? How negligible? Angle of attack, and hence vertical wind velocity 6 is likely to be significantly affected, but the horizontal wind may not depend on this very much. "although some differences in the frequency content of the vertical wind component could be expected." Why? What differences? Why is this important? Vertical wind velocity does not seem to be used later.

We have provided a much more detailed comparison of the substitution of the aircraft's angel of attack for the probe's angle of attack in the revised manuscript.

5. Generally, the details of this particular mutli-hole probe and its calibration and resulting accuracy were not provided. Can these be referenced from an earlier publication?

The probe used here is derived directly from the probes used in:

Barbieri, L. and Kral, S. T. and Bailey, S.C.C. and Frazier, A.E. and Jacob, J.D. and Reuder, J. and Brus, D. and Chilson, P.B. and Crick, C. and Detweiler, C. and others (2019) "Intercomparison of small unmanned aircraft system (sUAS) measurements for atmospheric science during the LAPSE-RATE campaign," Sensors 19(9), 2179.

and utilize calibration systems and approaches described in: Witte, B.M., Singler, R.F. and Bailey, S.C.C. (2017) "Development of an Unmanned Aerial Vehicle for the Measurement of Turbulence in the Atmospheric Boundary Layer," Atmosphere, 8(10), 195.

Al-Ghussain, L. and Bailey, S. C. C. (2022) "Uncrewed Aircraft System Measurements of Atmospheric Surface-Layer Structure During Morning Transition," Boundary Layer Meteorology, v185, 229-258.

We have added these references to the revised manuscript. Note, that these probes have also been successfully deployed in previous studies, including:

Bailey, S.C.C., Smith, S. W., Sama, M.P., Al-Ghussain, L. and de Boer, G. (2023) "Shallow katabatic flow in a complex valley: An observational case study leveraging uncrewed aircraft systems," Boundary Layer Meteorology, v186, 399–422.

Bailey, S.C.C., Sama, M.P., Canter, C.A., Pampolini, L.F, Lippay, Z.S., Schuyler, T.J., Hamilton, J.D., MacPhee, S.B., Rowe, I.S., Sanders, C.D., Smith, V.G., Vezzi, C.N., Wight, H.M., Hoagg, J.B., Guzman, M.I. and Suzanne Weaver Smith (2020) "University of Kentucky measurements of wind, temperature, pressure and humidity in support of LAPSE-RATE using multisite fixed-wing and rotorcraft unmanned aerial systems," Earth System Science Data, 12(3), 1759-1773.

Bailey S.C.C., Canter C.A., Sama M.P., Houston A.L. and Smith S.W. (2019) "Unmanned aerial vehicles reveal the impact of a total solar eclipse on the atmospheric surface layer" Proceedings of the Royal Society A, 47520190212.

so they, and their use, are not untested.

However, all these references relate to low altitude use. Has the probe been calibrated for use at the low pressures and temperatures encountered in this study? How do you know that the 20 Hz bandwidth (line 495) holds at high altitudes, since this depends on "viscous attenuation of the pressure fluctuations within the tubing" (line 252), and kinematic viscosity increases markedly at high altitudes.

As I am sure the referee is aware, conducting a wind tunnel directional calibration at the conditions experienced at high altitude would require specialized facilities that we did not have access to. We did verify operation of the equipment at low temperatures and pressures in an environment chamber prior to installation in the HiDRON (although we could not conduct velocity/directional calibrations in this chamber). However, we also repeated the calibration at the lowest velocity our tunnel can produce, as well as at conditions matching flight dynamic pressure, to check the Reynolds number independence of the calibration (as well as with and without the probe heat). However, prior to installation in the HiDRON we had not conducted the frequency response test in the environment chamber. This test was conducted since the last revision and the results are discussed in the revised manuscript.

6. A portion of the descent from the 30km release seems to be very steep. There were also some very tight circles at isolated points in the first two descents. Why?

   These flights were also test flights for the aircraft. During the flight, the operators conducted several tests of their systems which included adjusting the flight profile mid-flight and improving the response of the aircraft following release. Note that the steep release at 30 km is due to the requirement to achieve sufficient dynamic pressure to produce enough lift for controlled flight. The lower density means that the aircraft must fall a certain distance before the aircraft can travel to it's measurement location.

   I don't see any revisions to this effect. It would be interesting to know the "measurement ceiling" of this system.

   We have added a statement to this effect in our current revision. To our knowledge the aircraft's ceiling is not yet known, as the flights reported here were the highest altitudes the aircraft has been tested at to date.

7. central differencing between adjacent 30 sec averaged values?

   That is correct, although we have updated the revised manuscript to use spatially regular, rather than temporally regular, segments.

   As noted above, the effect on the record samples and time length would be helpful.

   We have included more rationale behind our selection of record length in the revised manuscript.

8. compare well given the spatial offset and the local weather conditions, and if the local periodic variations are ignored. These variations are suspiciously periodic with altitude, raising questions about artifacts from the platform airspeed/attitude/descent rate that may be varying with the same period. (See the related comments about Ri later). Some evidence should be provided that these results are not correlated with aircraft motions.

   This was addressed when responding to the initial comment. The additional radiosonde profiles in the revised manuscript also provide increased context for the wind fluctuations observed.

   The radiosonde wind magnitudes seem highly smoothed/decimated with straight lines connecting large changes in altitude. So much so that they don't really corroborate the "periodic" variations in the Hydron winds of concern. I would think that standard 1Hz cadence radiosondes would provide higher resolution.

   We agree. We tried to find the NWS data at standard 1Hz but could only find them with the resolution provided here, hence our use of the phrase 'publicly available' when referring to the NWS radiosonde data (as these data are likely to exist in a repository that we are unaware of). We also looked at ERA reanalysis data, but it had similar resolution to the radiosonde data we found. We note that the manuscript does contains a comparison to wind profiles measured during ascent by the aircraft's GPS that provide favorable comparison.

9. The Ri profiles seem to have highly periodic excursions with altitude. Could these be at the same period as the orbits the plane executes on the descent? That is, how do we know this is not an artifact of the sensors or the periodic motion of the platform? It would also be good to see how this correlates with the bank angle of the plane, since this will not be constant in wind. It will be difficult to take the results at face value without careful checking for motion/attitude/descent rate artifacts from the platform. Similarly, the Ri values seem suspiciously low, with < 0.25 values for much of the flight. Are these low values periodic anomalies in the measurements?

   We have extensively updated the text addressing this variability and present an improved method to calculate Richardson number. See additional discussion above regarding the dependence of wind velocity on the orbits.

   "The vertical gradient of each quantity was then found by finding the nearest statistical segments in the positive and negative z directions which minimized the difference in α". This is difficult to unpack (took me a while). I think you are saying that helix segments one turn apart, above or below each other, are selected for each alpha. This could be clarified better.

   This was want we were saying, we have simplified the description in the revised manuscript.

   Figure10: Rit is erratic and very small, Riz is less erratic but very large! As noted, this mismatch is due to the mismatch in Si vs Sz, since Nz essentially smooths Ni. But the reasoning behind the difference is S also applies to N, so the issue with S is "most likely" due to something else. Thus it is hard to believe either Ri value, or either S value.

   As can be observed by comparing existing profiles of wind and temperature, it is common for there to be more rapid changes in wind speed and direction than in temperature. This is also the case in the current study. We should therefore expect higher sensitivity of S^2 to the vertical scale used for calculating gradients than N^2. Hence the reasoning for these differences should not be expected to apply equivalently to N^2 and S^2.

   We have struggled with the sensitivity of Ri to the details of its calculation, but the method applied in the previous and current revision is similar to profiles presented in radiosonde literature. We have revised the discussion of N^2 and S^2 in the current revision to try to better reflect the sensitivity of the results to the method used for calculation.

10. $F_{uu}$and Φ_{uu} are both used for the power spectral density (please be consistent), and these are incorrectly called the "frequency spectrum".

   Please note that we are using colloquial language used in contemporary fundamental turbulence literature whereby the power spectral density is referred to as the velocity spectrum, or more broadly the energy spectrum, or shortened to simply the spectrum depending on context. Furthermore, these "spectra" are often distinguished as to whether their dependence is in the frequency domain or wavenumber domain by referring to them as the "frequency spectrum" or "wavenumber spectrum" respectively (see, for example, Pope (2000) "Turbulent Flows", Cambridge University Press.)

We made the distinction in the original manuscript between $F_{uu}(f)$ (i.e. the frequency spectrum) and $\Phi_{uu}(\kappa)$ (i.e. the wavenumber spectrum) as these are separate functions since both $\int_0^\infty F_{uu} df$ and $\int_0^\infty \Phi_{uu} d\kappa$ must both return the variance of the velocity component that they are calculated for (e.g. $\overline{u^2}$ in the above case). As $\kappa \approx 2\pi f/V_{rel}$ this means that $\Phi \approx F V_{rel}/(2\pi)$. We realize that this distinction is confounded by the fact that the original figure 5 was mislabeled as Φ_{uu} instead of F_{uu} for two of the subfigures, but this was a typographical error and not intentionally inconsistent. In the revised manuscript we strive to be more clear that we are referring to the velocity spectrum, and try to be more clear with the distinction between $F(f)$and $\Phi(\kappa)$. Reviewer 2 also did not like the use of $\Phi(\kappa)$, so we have now replaced that with $E(\kappa)$. We also now present the wavenumber spectra instead of the frequency spectra in the revised manuscript as it was more relevant for the discussion of determining the inertial subrange slope.

Because of the confusion in the literature, it would be helpful to distinguish power spectral density from amplitude spectra, power spectra, etc., since this makes a difference in the quantitative results. Note the Welch method can also be used for smoothing the amplitude spectrum, and the discrete Fourier transform can use different normalizations. To make it clear what was done, it would be good to include the expressions for the Fourier transforms used, or provide a specific reference.

The calculation was conducted using the Matlab function, pwelch, and have added that information to the revised manuscript. However, the documentation does not provide details of the analytical form of the calculation. That said, during this and previous studies we have extensively compared the output of the pwelch function to the frequency spectra calculated via FFT, and have confirmed that integration of the resulting one-sided frequency spectrum returns the variance of the signal (e.g. Saddoughi and Veeravalli, JFM, 1994, Pope, Turbulent Flows, 2000). The only difference we've noted is the additional smoothing due to the averaging windows, which we minimized by selection of window size in the present calculation (as also mentioned by the referee), which we found slightly reduced scatter in the corresponding results and was was why we elected to use the Matlab function rather than the results from the FFT calculation.

11. shape of the "spectra" in Figure 8 is very strange. Part a) seems to have a noise floor near $10^{(-3)}$, but the noise floor is smaller (near $10^{(-6)}$ at a higher altitude (part b)! (Where pressure fluctuations are necessarily smaller). A noise floor is again seen near $10^{(-3)}$ in part c). Also, the spectral slope is too shallow in part c) as noted later in the paper. Could it be that this power spectral density is the noise figure of the sensor itself, and there is really no detectable signal at these low atmospheric pressures?

As noted above we had mislabeled Fig. 8 due to changing from presenting the results in the wavenumber domain to the frequency domain. The original figures were actually presenting the results in the wavenunumber domain and we are unhappy that this error made its way into the final submitted manuscript. We also believe the reviewer is correct that the -1 slope is the noise figure of the sensor (or more accurately the combination of sensors used to determine the wind) and that the -1 slope indicates that there is no measurable turbulence present.

The noise floor noted above was a flat level, not a f^{-1} slope. This is typical of electronic Johnson noise in sensors. This can be seen at high frequencies in the new Figure 12 b) (at a level of about 10^{-3}) and c) (at a level of about 10^{-2}), which now makes more sense with the increase in altitude. Were the spectra in Fig. 12 cases 14 deemed Kolmogorov turbulence? If so, it would be good to show spectra for those that did not so the noise floor issue can be clarified against your fitting/qualification procedure.

It was not clear to us where you were seeing 'flat' energy roll off, particularly given the amount of scatter observed in spectra. Our -1 slope comment was more related to the fit to the range $0.1 < f < 20$ Hz) which was returning a -1 slope instead of a -5/3 slope when the energy content at high frequencies dropped below the noise floor (therefore reducing the average roll-off of the spectrum). However, this discussion is no longer relevant as the current revision has made several updates to the method used to find ε. Specifically, we now allow the fit range to vary by identifying the noise floor, and we fix the slope of the fit at -5/3 such that ε is now the only fitted parameter. We have also removed the individual spectra from the revised manuscript, replacing them with a wavelet transform of the entire flight time series.

12. Buoyancy Reynolds number can be calculated after estimating epsilon, as a check on this assumption.

This assumption is not expected to be valid over the entire measurement range, and is checked/unvalidated by the power law fit discussed on lines 456-457. However, assuming it's validity is necessary to produce a dissipation rate estimate using this approach. Validity of this assumption is therefore reflected by the filled circles on Figures 9 and 10.

Anomalous spectral shapes can return a $f^{-5/3}$ fit. It's the standard deviation of fit that is telling, and this not reported. Lines 456-457 in the revised manuscript concern N and S, not epsilon. The case of flight 1 in figure 12 is strange in that it has EDR values higher than many points in flights 2 and 3, yet these values are not qualified as turbulent. Also seems strange that the EDR is higher but TKE is smaller for flight 1. This does not make sense. How do the spectral shapes compare? Seems the sensor direction in flight 1 has been fixed from the original paper, but are there any other sensor system differences between flights? Did w' make a significant difference in the TKE calculations for flights 2 and 3, given the comments earlier that w' is likely to be incorrect?

We agree with the reference to anomalous spectral shapes and value of R^2 for fit accuracy. The lines referenced were the prior submission and hence why they weren't relevant to the revision. As noted above, we have adjusted our ε calculation to integrate only to the noise floor, which has addressed many of the concerns noted above. With regards to w', as shown in Appendix A in the revised manuscript, the impact on w' of replacing the disconnected transducer was not as large as could be expected and did not make a significant impact on the TKE calculation.

13. Not clear how (of if) the noise floor/noise figure is removed in the qualified data fits.

We made no attempt to remove the noise floor from the data fits.

The noise floor can corrupt the estimated slope, so its removal can often retrieve epsilon values that were previously rejected. This could help to explain the absence of qualified values at the higher altitudes (ala line 552),

In the latest revision we have removed the noise floor from the data fits. This process is explained in the revised manuscript. We found that it did 'clean up' the profiles of ε a little bit.

14. confirmation that the infrasound signal is due to turbulence is too strong a conclusion at this stage. Localized increases do not correspond to those in EDR.

This statement was driven by the increase observed as the aircraft enters the boundary layer and becomes immersed in the turbulence (as evident in the comparison of Figure 10 and Fig.12) and also the comparisons in Figures 14-16. However, we agree that this conclusion is weak, and that the infrasonic measurements need more development.

570: "with the range used selected due to finding that lower frequency acoustic content better correlated with the EDR values measured with the five-hole-probe when compared to the higher frequency acoustic content, which tended to contain additional signal noise."

This is problematic: seems like you are cherry picking aspects of the infrasound signal that correlate with EDR, in order to confirm that infrasound signals 16 indicate turbulence (EDR). What is the nature of the "additional signal noise"? 577: despite agreeing that "this conclusion is weak" above, the line "providing an initial confirmation of the presence of infrasonic sound generation by turbulence" is still present in the paper.

We have revised the infrasonic microphone analysis and find that normalizing the low-pass filtered microphone response by the unfiltered response we are able to get good agreement between the long wavelength wind velocity variance and the infrasonic (sub 20 Hz) acoustic energy. This threshold was selected to conform with the typical definition of infrasonic frequencies.

15. 592: "We therefore also only examine the behavior of other quantities in terms of relative trends in these values, rather than specific thresholds." Vague.

This section is no longer included in the revised manuscript.

16. 595: "Examining Flight 1, the distribution of measured Nz shown in Fig. 14a indicates that the static stability conditions are nearly constant with altitude." Stability is significantly increasing with altitude! This was also stated earlier in line 469.

This was a typo and should have read constant with aziumuth. Regardless, this section is no longer included in the revised manuscript.

17. 655: “were actually produced by the aircraft passing through the same structure more than once.” This is not clear from the discussion in that section. Pointing out a specific example or two would help to support this comment.

    This section is no longer included in the revised manuscript.

18. 662: Despite this ambiguity, these initial flights suggest that the sUAS measurements suggest the potential exists” Please reword.

    The Summary and Conclusions section has been rewritten.

19. 666: “the flight pattern allows for increased statistical convergence due to the larger volume of air sampled over a particular altitude range.” How does a larger sample volume automatically provide “increased statistical convergence”? This is likely to produce more variation in measured quantities. Also, the conclusion is not a natural place to introduce a specific argument.

    The Summary and Conclusions section has been rewritten.