Referee 3 Response to Manuscript Revisions

Referee comments on the original manuscript are in black. Author responses are in blue. Referee responses to the revised manuscript and author responses are in green. (Apologies for the rough formatting of space).

We appreciate the time taken by all referees in providing insightful and detailed comments about our manuscript. Following the reviewers recommendations, numerous changes have been made in the analysis approach including: (1) Revisiting the conversion of the five-hole-probe wind estimate; (2) adjusting the statistical ensemble sized from being based on 30 s ensembles in time to 1500 m ensembles in space (with 750 m overlap between ensembles to retain spatial resolution); (3) revised the spectra calculation with the correct transformation from frequency to time domain; (4) updated the method used to calculate vertical gradients and adding Brunt-Va¨isa¨la¨ and shear frequency measurements to the paper; and (5) addressing issues in the methodology used to generate contour plots. Of these changes, the largest impact on the results was the change made to the vertical gradient calculation, which impacted the gradient Richardson number, Ri, values and the revision to the contour plots. As a result of these changes, and other additional changes made to address specific comments made by the reviewers, we believe the revised version of the manuscript is more clear and complete than the version originally submitted. To help clarify where changes have been made in the revised version, we have highlighted all changes made using blue text.

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

Reviewer(s)' Comments to Author:
This paper presents some intriguing results using a new measurement platform for profiling the atmospheric column descending from about 20km. However, the quality of the observed data is not clear, given the periodic variations that may be a result of the periodic orbit of the gliding aircraft platform. As a result, the conclusions drawn about the viability of the sensing method and the relation to potential atmospheric structures and sources is tenuous, without furhter examination of the correlations between signal variations and platform motions.

As we were also concerned with apparent wind magnitude variations with altitude having a vertical wavelength similar to the altitude difference between successive aircraft orbits, we have thoroughly revisited the wind measurements and found several places where improvements could be implemented, including improving the time alignment between autopilot kinematic variables and payload sensors, identifying and correcting pitch and yaw probe misalignment (of less than 10°) between aircraft and sensor coordinates, and discovering an error in probe rotation. However, these improvements only marginally modified the wind magnitude and did not affect the observed vertical profile in any meaningful manner.

We have also closely examined the dependence of the vertical profiles of wind magnitude with heading, as shown in Figure 1. The most notable similarity between vertical separation of orbits and vertical wavelengths in wind magnitude occurred during Flight 1. However examining successive orbits shows that the wavelengths are not identical, with the orbit vertical distance slightly longer. Therefore, if there is a bias in the measurement, it is not rigidly correlated to the heading.

We also perturbed different inputs into the wind estimation, and found that the vertical wavelengths in the wind estimate appears to be robust to these changes. In summary, we could find no conclusive evidence of bias in the measurements introduced by the aircraft heading and the periodic orbit.

Detailed comments:

Introduction: "The trajectory of the glider allowed for improved statistical convergence and higher spatial resolution of derived statistics measured by the in-situ sensors." Refers to balloon-borne measurements, but such improvements and higher spatial resolution were not demonstrated in the paper.

Here we were referring to the airspeed of the glider able to transect the flow at flight speeds over 20 m/s which, when compared to balloon-borne measurements, means that more wavelengths of turbulence can be measured over the same duration of sampling time. However, spatial resolution has different connotations and is not the correct term for what we are trying to describe. We have revised the text to better reflect our intended meaning.

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”.

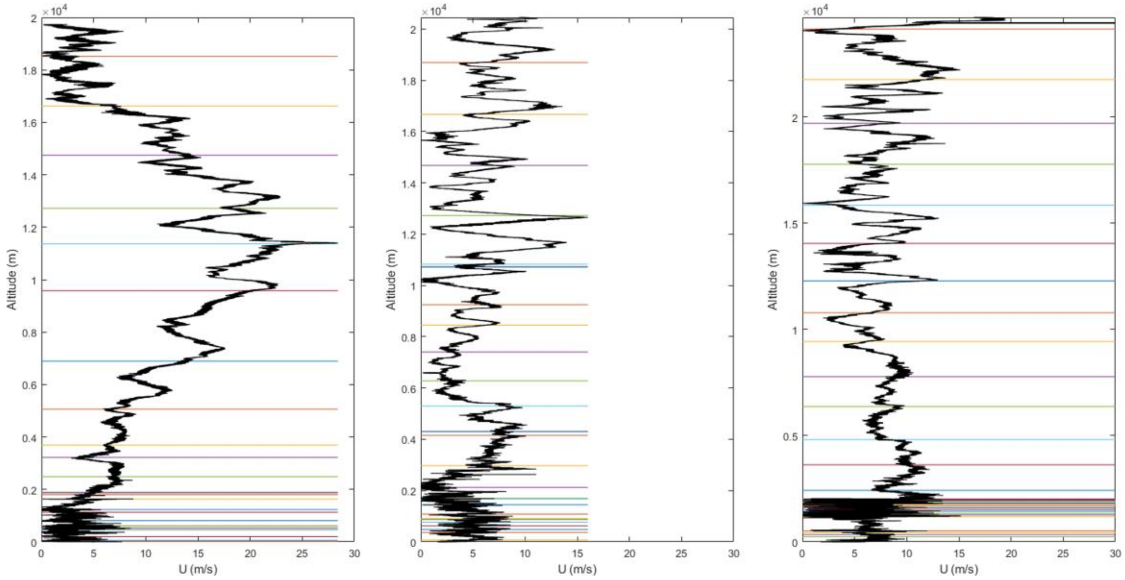

Figure 1: Vertical profiles of wind magnitude from all three flights shown with horizontal lines indicating location where aircraft heading passed through 180˚. (left) Flight 1 (middle) Flight 2 (right) Flight 3

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

The abstract still contains this same phrase. 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.

40: "with these results used to model the relationship between turbulence in the stratosphere as well as tropospheric activity" not clear: "as well as" vs. "and"?

Revised. 'and' should have been used.

85 "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.

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).

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 5m/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.

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

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. 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.

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. 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.

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.

125: The iMET sensor specifications were not referenced. These accuracies and time constants tend to degrade at lower pressures and temperatures, and this was not indicated.

We have added a citation to the manufacturer's webpage where this information was obtained. Note that these sensors were previously flown on a similar platform and found to be consistent with model results (Schuyler TJ, Gohari SMI, Pundsack G, Berchoff D, Guzman MI. Using a Balloon-Launched Unmanned Glider to Validate Real-Time WRF Modeling. Sensors. 2019; 19(8):1914. https://doi.org/10.3390/s19081914).

Figure 3 would benefit from the addition of dimensions to the components pictured.

Dimensions have been added to the figures

150: "Comparison of calibrations with and without heating active indicated that there was no influence of probe heating on the five-hole-probe response characteristics." Not clear what response characteristics means: time constant? Calibration coefficients? Noise level?

We are referring to the calibration surfaces and have revised the text to be more precise.

153: "Each hole on the probe was connected to differential pressure transducers through 1.75 mm diameter flexible polymer tubing." What was the other port of the differential pressure sensor connected to? Presumably this was the "static pressure port", but this was not shown or described in the paper. How long was the tubing (this can have a detrimental effect frequency response of the air speed measurement, as noted later in the paper).

We have added these details to the text.

160: "Note that the during flight, the autopilot maintained flight speeds sufficient to produce pressure differences well within the range of the low-sensitivity transducers and hence only the readings from these sensors were used for this analysis." Please quantify the airspeeds obtained, and the corresponding average differential pressures.

We have added figures showing relative air velocities and noted the value of dynamic pressure in the text.

195: 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 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.

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.

205: How was the microphone mounted on the vehicle? Was it protected from dynamic pressure fluctuations? If so, how did this filter the infrasound pressure waves? Could aircraft motions (that are also dependent on ambient turbulence) influence these measurements?

The microphone was mounted rigidly within the nose of the aircraft, with the diaphragm facing forward. Being within the fuselage, the microphone was protected by dynamic pressure fluctuations. Note that infrasonic sound waves will be of the order of 30 m and larger, so attenuation of the sound waves by the fuselage is not expected in this configuration. Due to the rigid mounting of the probe in the aircraft, it is not anticipated that aircraft motion could influence the microphone, however we were not able to verify this assumption from the current set of measurements.

We have added this information to the revised manuscript.

How are winds calculated?

Winds were calculated based on the procedures described in lines 153 to 187 (lines 199 to 231 in the revised manuscript).

How is airspeed calculated? No plots of airspeed were provided. What was the airspeed as a function of altitude?

Airspeed was measured by both the aircraft's Pitot probe and the five-hole probe using the standard procedure of measuring dynamic pressure across total pressure (central hole) and static pressure and found to be in agreement between the two instruments. Plots of relative air velocity have been added to the revised manuscript.

220: temporal alignment can be intricate. How was this accomplished with this data? Was there a common time reference?

Additional clarification added to text. Initially we intended to use some of the statistics calculated by the payload and sent to autopilot via RS232 communication and recorded in the the aircraft telemetry stream, but found correlating the dynamic pressure measured from the aircraft pitot probe and measured from five hole probe to be a more reliable alignment indicator due to suspected buffering delays in the RS232 connection. Note that, although not mentioned in the manuscript for brevity, we were able to identify and remove a 0.005% difference in clocks between the two systems by windowing the correlation of the airspeed and five-hole probe. This corresponds to a 1 second difference in timing over the six hours of measurement.

230: what does "controlled landing" mean here? Manual landing (RC), or automatic landing (autopilot)?

Landings were conducted by the autopilot. This has been noted in the revised manuscript.

229: 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.

235: I think you mean UTC -6:00 here.

Correct. Revision made.

239: "Due to the configuration of the sensors on the aircraft". Vague. Please describe what about the configuration makes the sensor readings unreliable on ascent.

During ascent, they are in a stagnant region within the wing pillar wake and therefore not sufficiently aspirated to prevent self-heating and delayed air exchange with the environment. The result was that the T and RH measurements had a significant altitude-dependent warm bias of 5 to 15 degrees K relative to descent and the radiosonde measurements. On descent, these sensors are introduced into the oncoming airflow and properly aspirated. We added more details to the revised manuscript.

289: "with backing"?

Revised.

295: 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.

230: regression fit to a constant function over 150 sec? Central 30 sec interval with 2 intervals before and 2 after?

Prior to differencing we employ a 5 point moving average where the smoothed value at the central segment is found by averaging the values averaged within the central segment with average values from the 2 segments before and 2 segments after. We have added more details in the text.

271: "this is likely due to spatial heterogeneity in the atmospheric moisture concentration". Could also be due to an instrumentation anomaly.

It is possible. We have added more radiosonde data to these figures to better represent the spatial heterogeneity. The results suggest there might be some dry bias in the RH sensor in cold temperatures.

The following statement "cloud conditions near Truth or Consequences, NM (near Spaceport America) were different" does not help. Different how? At what altitudes?

The ASOS reports only qualitative conditions (i.e. scattered, clear, overcast, etc.). However, the additional radiosonde profiles made the ASOS reported cloud unnecessary and this text was removed. We also now include satellite imagery in Appendix A, which provides a more nuanced illustration of the different humidity conditions which could be expected.

276: 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.

300: Although it probably does not make much difference, z in this formula should be altitude MSL, not AGL. Recommend that MSL be used throughout for consistency and for interpretation of the results. Also, I can't seem to find the altitude of the ground at the launch location.

True, this was a typo as we made our calculation using MSL. However, based off of Reviewer 2's comments, we have removed the altitude dependent g calculation from the revised manuscript. We have also added the m.s.l. altitudes of the launch and recovery location manuscript.

303: 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.

455: "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.

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.

323: how does $< u >$ differ from $< U >$ used earlier?

We use u, v, and w to indicate the components of the wind vector having a projected horizontal magnitude of

U. These distinctions were defined on line 183 in the original submission with U specifically defined in equation 3 (now equation 7 in the revised manuscript) .

327: what were the subintervals and overlap used in the Welch method? The “spectra” in Figure 8 seem to have a 40s period fundamental frequency, so this is confusing.

We applied the Welch method without using only three subintervals and a 50% overlap. We have also verified that the application of Welch’s method does not impact the shape of the resulting velocity spectrum. The 40s fundamental period is due to mislabeling of the figure, as discussed later.

F_{uu} and Phi_{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 $\int_0^\infty F\,df$ and $\int_0^\infty \Phi\, d\kappa$ must both return the variance of the velocity component that they are calculated for (e.g. u 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.

328: Was the Hanning window variance-preserving?

Yes. We confirmed it is variance preserving prior to implementing it.

329: integration of the PSD is from .025 Hz to 5Hz, so the most important (largest amplitude) components of TKE are potentially not included. This makes TKE difficult to estimate without some idea of the local “outer scale” where the spectral energy ceases to increase as frequency decreases. This is noted later (347), but with a confusing reference to 300m as the longest period in the k “measurement”, since earlier in the paper 2420m was quoted as the longest interval in the 30s analysis intervals. However, no mention of the outer scale was made. Thus the k retrieval has a highly variable lower spatial scale with altitude, and the relevance of the k profiles is unclear.

We had used 300 m as it is the average distance travelled during the 30 s sample duration, but acknowledge that the airspeed at high altitude was much higher, leading to the 2420 m length of the longest segments. In the revised manuscript we use a constant statistical segment length. Although this may not fully resolve the largest eddies it should provide a consistent wavenumber range for which TKE is calculated. We did work at trying to implement an approach which spatially varied the segment length depending on the energy content, but due to the range of scales experienced during descent, we were not happy with the implementation and feel it needs further development before including in a publication.

339: 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 waveunumber 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 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.

356: 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?

364: the κ_1 wavenumber component in (9) is the longitudinal component of the motion of the air relative to the sensor. This is only the longitudinal component of the vehicle ground velocity in the special case of zero mean wind (still air), or in the limit when the vehicle airspeed is much larger than the wind speed. This should be corrected to use the airspeed, and the course heading frame rotation should be replaced by one based on angle of attack and sideslip of the sensor relative to the relative wind vector.
366: again use of ground speed here in incorrect. Must be airspeed.

Thank you for raising this concern. Our initial approach was selected in a misguided attempt to minimize the effect of bias introduced by the dependence of Taylor's hypothesis on the wavenumber dependence of the wind velocity used for its application (discussed by Moin, JFM, 2009). Our experience in the ABL has indicated that neglecting the advection due to mean wind does not impact the resulting spectra, however these measurements are typically made in winds an order of magnitude smaller than the UAS ground speed. As pointed out by the reviewer, this is not necessarily the case for the measurements reported here. We therefore have revised the calculation of the longitudinal wavenumber spectrum in the current version of the manuscript to account for the advection due to mean wind. This includes adjusting the estimate of wavenumber from aircraft relative air speed and using the relative velocity component of the wind when calculating the spectra.

367: I don't understand the expression for $\Phi(\kappa_1)$.

This is discussed above and relates to the property that $\int_0^\infty F\,df$ and $\int_0^\infty \Phi\,d\kappa$ must both return the variance of the velocity component that they are calculated. We have updated the text to provide a better indicator.

Figure 9: Suggest plotting k and epsilon (or EDR) on a log scale to look for periodic artifacts (as noted earlier), and to make it easier to see the full range of these power function values.

We have plotted these figures on logarithmic axes in the revised manuscript and find no clear indication of artifacts due to athe periodic orbits.

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),

386: "It will be shown later that this enhanced EDR corresponds to measured fluctuations in velocity introduced by the presence of gravity waves at these altitudes". How? Gravity waves have a much large wavelength than could be influencing these epsilon estimates.

Unfortunately, the appearance of gravity waves appears to have been contaminated by the two-dimensional interpolation technique used to in the $\langle T\rangle'$ and $\langle w\rangle'$ analysis used to discern the presence of gravity waves. Although we believe there is still merit in the concept behind the approach, we could not come up with a suitable implementation in time to include in the revision. We have therefore removed the gravity wave discussion from the revised manuscript.

401: 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 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.

426: the "interesting features" in figures 12-14 show strong correlation with location on the flight path circle. This might be due to differences in structure in the atmosphere across the 5km circle diameter, but it might also be due to sensor signal dependence on the heading or attitude or airspeed of the vehicle, that is also periodic with location on the flight path (as noted above). Given that these "features" in the data persist over large altitude ranges (where e.g. shear and stability are expected to vary significantly), and the intermittent, sometimes contradictory correlations noted in the paper, it is difficult to consider the conclusions offered as more than optimistic interpretations of rather murky relationships. Too much is made of data that has not been thoroughly vetted.

We note that the statistics being presented are calculated over relatively short windows (of a length about 5% of the orbit circumference) and, as many of them require mean subtraction, they should be largely unaffected by any bias which may be introduced by aircraft heading.

When preparing Section 3.4 we were also concerned with the potential for bias due to interpolation and had ensured that there was at least one measurement point per interpolation grid cell. However, when revisiting these figures during the revision we found that the numerical interpolation scheme (which was an implementation of Delaunay triangulation) was creating cell values very different from the measurement point within the cell. Therefore, although not appearing to be due to position in the orbit, the reviewer's concerns were well-founded.

As Section 3.4 interpreted these figures in the context of $\langle \phi \rangle'$, and this approach used the interpolated data for background subtraction, we have removed this approach from the manuscript completely rewritten this section (which is now Section 3.5 in the revised manuscript). We still believe that there is merit in this visualization approach, however, since it does show interrelationship between statistical values determined at measurement points at the same azimuthal locations along the orbit. In the updated section, the contours are determined by direct triangulation between points and we have included the actual measurement locations and values in the revised figures to allow the reader to directly assess the impact of the triangulation on the contours.

449: Ri is used as a marker for stability in various places, which is confusing. Stability is indicated by N. Ri combines N with horizontal shear.

We have revised the manuscript by inclusion of square Brunt-Va¨isa¨la¨ frequency N^2 and square shear frequency S^2 as they better reflect our intent.

441: an "identification" the source of observed EDR here is optimistic here, given the weakness of the "suggestions" seen in the data. Again, "wave" activity may be due to measurement anomalies that are periodic with vehicle motion.
451: given the strong horizontal advection, it is difficult to believe that turbulence features originating in the boundary layer could propagate into the stratosphere within the short 5km diameter of the (inertially fixed) helix of measurements.

As noted above, the waves and some of the vertical features are likely to have been introduced by the numerics of the interpolation scheme used and these interpretations have been removed. We have significantly updated section 3.4 and feel the the inclusion of S^2 and N^2 distributions make the connection between EDR and shear/buoyancy easier to infer.

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

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.

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.

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

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.