Response to Referee 1

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-Väisälä 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:

Haghighi et al. present an exciting new measurement system that allows to sample the atmosphere up to 25 km with high resolution. It is great to see that the authors could perform these measurements and that they could measure exciting features in the atmosphere. I have some concerns about data processing and analyses which need to be revised before I can recommend the manuscript to be published in AMT.

General comments:

- I do not trust the Ri-number calculations which are presented. The oscillations that can be seen in the vertical profile seem unrealistic and I have a strong suspicion that the error starts with the wind measurement, as described below in the specific comments. Please check and make sure that the wind measurements are not heading-dependent and estimate the uncertainties for your system.

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 modified the wind magnitude by 10% at most 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.

Note that a formal error propagation analysis is quite challenging for this type of measurement, as it involves numerous sensors on the aircraft autopilot (GPS, gyroscope, accelerometers, and associated Kalman filters, timing clock), on the payload (pressure transducers, thermistor, capacitive hygrometer, data acquisition systems, and timing clock) and on the wind tunnel calibration (Pitot probe, directional gimbal accuracy, data acquisition system). Therefore providing a formal uncertainty estimate is non-trivial. We therefore utilize the less formal approach by perturbing different elements of the calculation, and found that the wind estimate appears to be robust to most of these changes. Thus we can only roughly estimate our uncertainty using previous intercomparison studies with a very similar probe at ± 1 m s^{-1}.

Finally, we note that the Ri periodicity is directly related to the calculation approach, whereby we simply

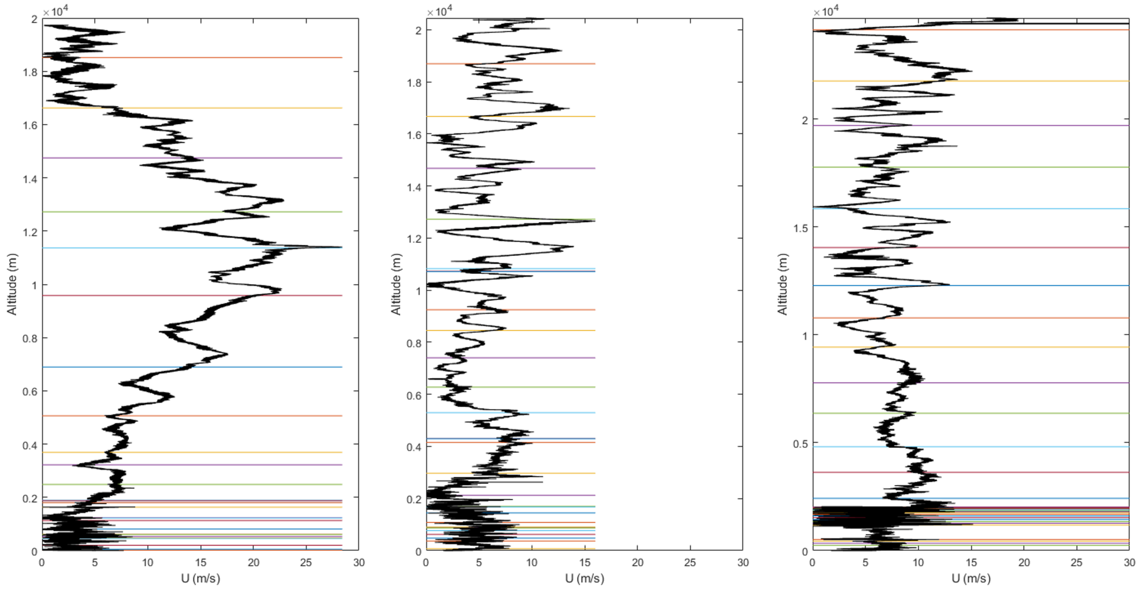

Figure 1: Vertical profiles of wind magnitude from all three flights shown with horizontal lines indicating location where aircraft heading passed through 180°. (left) Flight 1 (middle) Flight 2 (right) Flight 3

calculated the vertical gradients from the time dependence with of altitude. However, this approach assumes horizontal homogeneity in properties along the aircraft orbit, which may not be strictly correct over the distances travelled, particularly with the wind measurements. Correspondingly, these values were also sensitive to the ensemble size used. We therefore present an updated approach within the revised manuscript that attempts to calculate Ri using a more localized vertical gradient.

- The title and the abstract suggest a deeper analysis of the infrasound measurements, the error sources and the uncertainties, but the manuscript leaves me with the impression that these measurements can not really be used because it is often unclear where the detected signals originate from and how to process / correct them. It would be good to at least show a path forward for this.

We have expended significant effort examining the infrasound measurements, and your assessment is correct that although we gained understanding of the system and signals, we have only been able to identify qualitative relationships between the infrasound amplitude and the presence of turbulence. Ultimately, we can attribute this to two reasons: (1) the turbulence above the boundary layer that occurred during these measurements was insufficient to generate a signal significant enough to be sensed; and (2) we were unable to find a scaling that successfully removed the effects of density on the microphone response for all three flights. However, the increase in amplitude within the boundary layer and at some localized regions within the troposphere do indicate that the sensor was detecting turbulent signals. We believe that additional measurements with some planned modifications to the system will allow greater insight into how best to deploy the sensor operationally.

- The gravity wave analyses are very nice. The polar plots give some insight, but it would be easy to analyse the origin of the waves better when some model data is included and 2D maps of the wave structure at different altitudes was shown. I encourage the authors to consider that to make their very vague and speculative statements more robust.

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.

Specific comments:

p.8, l.192: was the attenuation of the probing also verified by experiment? what can you read from the spectra?

The attenuation was measured directly by introducing a step pressure change at the probe tip and measuring the voltage response. The settling time following the step change was 0.06 s which corresponds to a frequency response of 20 Hz. The velocity spectra only show the tubing influence for $f > 100$ Hz, however we use 20 Hz in the revised manuscript as a conservative estimate of the highest frequency accurately captured.

p.8, l.195f: if this data is shown, you need to explain the method in more detail.

We have expanded on this discussion further in the revised manuscript.

p.11, l.253: Where do these 10% come from? What do they even mean? It does not look like there is a relative error below 10% for wind speed and wind direction at all times. I would not expect it, given the temporal and spatial separation, but the value should be explained.

We had been referring to an average difference. However, in response to reviewer 2's concerns about the radiosonde comparison, we have revised this section through inclusion of additional radiosonde measurements at different locations and updated the text accordingly.

p.14, Fig.7: The periodicity in the Ri-number with height is very suspicious. Looking at Fig.4 I get the feeling that this could be caused by flight direction. It is known that wind estimatino from multi-hole probes on fixed-wind aircraft is very sensitive to heading estimation. Please show the dependency of your wind measurements to heading. I think the periodicity already shows in wind speed and wind direction measurements. I doubt that the Ri-number calculations are meaningful with this uncertainty in the wind estimation. It should be qualified with an uncertainty estimation.

This comment has been discussed and, we believe addressed, above.

p.18, l.341: how were the thresholds for n chosen here?

This range of values were somewhat arbitrarily selected as being within $-5/3 \pm 10\%$. We have added this range to the revised manuscript.

p.18, l.362: so, if I understand this correctly, in a circular flight, the horizontal flight direction changes all the time and thus does the wind vector component you are using for EDR estimation. I think it would be more reasonable to align the rotated wind vector to the mean wind direction. u and v are not expected to show the same spectral characteristics. In a circular flight you are also distorting the measurement, even if the Taylor hypothesis is valid. Maybe the radius is so large that within 30 seconds, the curvature can be neglected, but you should reflect on this.

The inertial subrange scaling described by equation 9 (equation 11 in the revised manuscript) is only valid in the longitudinal direction, i.e. specifically for the component of the velocity vector aligned with the component of the wavenumber vector. As the only wavenumber vector component we are capable of measuring is along aircraft's flight trajectory, this is the component of the velocity vector that should be included in the calculation of the energy spectrum. Although the mean value of u and v velocity components may not be equal, we subtract the mean value during the calculation, and therefore are only calculating EDR using the velocity fluctuations in those directions (which should be homogeneous and isotropic in the inertial subrange assumed by equation 11 and therefore not impacted by direction of flight relative to the mean wind). Note that in our original calculation, we assumed the advection of the eddies due to the mean wind was negilible, which the other reviewers pointed out may bias the results. Both our wavenumber and velocity component used to calculate the spectra have been revised to adjust for the advection velocity and the manuscript has been updated with revised example spectra, ε and EDR values.

p.20, l.414: This is a bit misleading. You did not add horizontal flight legs at these altitudes, it is still the same flight pattern, spiraling down, right?

This is correct. This statement was only intended as a generality based on the aircraft descending only 2 km for every 30 km of horizontal flight along the orbit. We have revised this statement to be more precise.

p.21, l.428ff: These are quite interesting observations. It would help to show the temperature and velocity fluctuations for distinct altitudes on a horizontal (map) plot.

As noted above, we found that these fluctuations were contaminated by the interpolation scheme used and we have removed the discussion of these fluctuations in the revised manuscript.

p.21, l.429: How do you determine the wavelength?

Wavelength was determined by $d\lambda = R d\alpha$ where R is the radius of the aircraft orbit and $d\alpha$ the distance between peaks in the periodic observations.

p.22, Fig.12: The shaded region is Ri¿1 or Ri¡1? It is not clear from the caption.

the shaded region was intended to indicate indicates more stable conditions ($Ri > 1$). The revised Ri calculation approach however has resulted in Ri values much higher than unity, so we do not include this region in the revised manuscript.

p.25, l.465: you mean figures 12 d and e, right?

Correct. We have fixed this typo.

Figs. 12, 13, 14: The variable names and units should be given in the plot itself, not only in the caption.

We have added variable names and units to the figures.

p.27, l.534: I would highly recommend to obtain some reanalysis data from NWP models (e.g. ERA5) to see if conditions were favourable for gravity waves and if they can be seen in the model. This could be nicely added in an appendix.

We did contact someone familiar with satallite observations and investigated reanalysis data as part of the revision process. However, as noted above, the gravity wave discussion has been removed from the revised manuscript due to a lack of confidence in the initial results.

Technical corrections:

p.1, l.3: "thermodaynamic"

Correction made.

p.3, l.61: "from with"

Correction made.

p.7, l143f: two times "changes with the horizontal axis" should probably be vertical axis the second time.

Correction made.

p.8, Eq.4: *dir* is not a proper variable symbol.

We have changed *dir* to γ/

p.8, l.194: disconnected disconnected

Correction made.

Response to Referee 2

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-Väisälä 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:

Review of the manuscript "High-altitude atmospheric turbulence and infrasound measurements using a balloon-launched small unscrewed aircraft system" by A. N. Haghighi et al.

General comments:

The authors present an exciting UAV system that mainly includes a five-hole probe and an infrasonic microphone for probing turbulence in the troposphere and lower stratosphere. The technical developments are unquestionably to be welcomed. They may represent a new step towards a technology better suited to in situ turbulence and small- scale structure measurements at high altitude, particularly in the stratosphere, by providing decisive information on the coupling between the fine-scale stratification, mixing processes and gravity waves. However, the proposed article appears to have major shortcomings that should be remedied before a possible publication.

If some of the issues are due to misinterpretation, the description of the methods should be clarified. There are also a number of inaccuracies or blunders which should also be corrected and some parts should be expanded to facilitate the interpretation of the results. Consequently, in view of the potential interest of this work, I propose that it be publishable after a very thorough revision and perhaps re-evaluation of certain parameters. The review does not comment the part relative to infrasonic measurements because I don't have enough knowledge to evaluate it.

Major comments:

(1) The proposed comparisons between temperature, humidity and wind profiles from radiosondes and sUAS cannot be conclusive, as they are made with data that are separated by several hours (up to around 6-7 hours) and launch sites separated by 160 km (this information only appears on line 270 when comparisons are discussed), and the fields are not stationary. Under these conditions, performance evaluation is difficult, if not impossible, and cannot "allow validation", as stated line 248, because disagreements can always be explained by the non-colocalization and non-simultaneity of the measurements. It would have been more useful to include a radiosonde under the balloon during its ascent, in order to make more appropriate wind comparisons. The PTU IMET-XF data during the ascent may provide better conditions for temperature and humidity comparisons, even if they are corrupted at small scales. On the other hand, the representation of data points with circles, rather than continuous lines, systematically used in the figures, does not allow simple comparisons between profiles and identification of the peaks referred to in the text. From a qualitative point of view, the three wind profiles measured by sUAS seem to indicate fluctuations compatible with the presence of gravity waves at all altitudes with a vertical wavelength of the order of 1 km or less, whereas the balloon data do not seem to reveal such fluctuations (but, once again, the graphical representation makes analysis difficult). The technical data of the

radiosondes used (such as the vertical resolution) should be indicated. Current standard sondes are now able to measure profiles at 1 Hz, which does not seem to be the case here. We can wonder whether the wind fluctuations observed by the sUAS, and apparently not by the balloon, are the result of an instrumental artifact or not. A spectral analysis of the whole (sUAS and balloon) wind (and T) profiles would provide useful information to check the consistency between the data.

We agree that the radiosonde comparison is not conclusive and tried to reflect that only a broad general agreement was present in the text. Unfortunately, at the time of the test flights, we did not have the capability to include radiosonde measurements in the test campaign and instead, as noted in the paper, used publically-available National Weather Service (NWS) soundings for comparison. As such, the resolution of these measurements were likely downsampled prior to being made available and we do not have access to the higher resolution original data set.

In the revised manuscript, we have added more information about the NWS radiosondes, and have added additional profiles made from two other weather stations nearest to the experiment site. We believe these additional soundings add context to the differences observed between the aircraft and weather balloon measurements.

We have added technical info about the NWS radiosondes to the manuscript.

With regards to using the iMet-XF data on ascent, we intentionally discard it as we found there to be non-negligible differences between the ascent and descent which are caused by the sensors placement. During ascent, they are in a stagnant region within the fuselage wake and therefore not sufficiently aspirated to prevent self-heating and delayed air exchange with the environment. They are also close enough to the fuselage to be within the thermal wake of the aircraft. The result was that the T and RH measurements had a significant altitude-dependent warm bias of 5 to 15 degrees K on ascent relative to descent and the radiosonde measurements. On descent, these sensors are introduced into the oncoming airflow and properly aspirated, resulting in better comparison with the radiosonde measurements. Note that the Hidron H1 used a similar setup and found good agreement with WRF model results on the descent (they also found significant difference between ascent and descent: 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

With regards to the formatting of the profiles using symbols instead of lines, this is simply an artifact of a preference to use symbols to present discretely acquired data points wherever possible to avoid biasing the viewer regarding the interpolation between these points that comes with connecting the discrete data points. However, in the revised manuscript we have replotted all data presented in profiles using lines instead of data points.

In summary, although we now wish that we had conducted co-located radiosonde launches during these flight tests, we neglected to do so and have to resort to publicly available data for comparison. We've updated the comparison with additional soundings in the revised manuscript, but providing co-located data is not feasible without repeating the experiments.

(2) The calculation of ε described on page 18 contains several important errors that need to be corrected. The conversion of the frequency spectra into wavenumber spectra must be made with the magnitude of the relative air wind speed, NOT the aircraft's ground speed. Description of the method can be found in Frehlich et al. (JAS, 60, 2487-2495, 2003) or Kantha et al. (PEPS, 4:19, 2017). As the difference can be important, especially at high wind speeds, the impact on ε should be far from negligible. As a corollary, the longest wavelength of velocity fluctuation is not given by the sole horizontal velocity of the aircraft, but by the relative velocity (lines 348-350).

Because the relative air wind speed may significantly vary with altitude (which is why it should be quantified and shown), TKE estimated for frequencies $f < 5Hz$ and shown in Figure 9 is likely not correct because it indicates the total energy for variable wavenumber bands. In addition, k_1 as defined in (9) is NOT the component of the wavenumber vector in the direction of the flight path (line 362), but along the direction of the relative wind vector formed by the direction of flight path and horizontal wind vector. There is no difference between the two, only when the UAV flies in the direction of the wind. Therefore, TKE dissipation rates (and TKE) must be recalculated.

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.

As for the frequency cutoff used for integration and its dependence on the relative velocity of the aircraft to the air, we note that our application of this cutoff in the frequency domain is directly analogous to any measurements made with any sensor having finite frequency response (i.e. any sensor measuring in time). However, to address the reviewer's concerns we have compared the TKE and Reynolds stresses calculated with a finite frequency cutoff and a finite wavenumber cutoff and found no difference. This is because the contribution to overall variance at the higher frequency/wavenumber ranges measured by the sensor is minimal so small differences in integration ranges at the high end have no difference.

What did make a difference is that, as noted above, we have modified our statistical approach to use 1500 m long ensembles rather than 30 s long ensembles. This change impacted the low wavenumber bound of the variance calculation and this change did modify many of the statistics (most notably TKE, as could be expected).

(3) Section 3.4

The method used to reconstruct the distributions of parameters in the $\alpha - z$ plane is not clear. The authors should very clearly explain how the interpolation method works as this is not a standard method of visualizing the data. However, it does not seem to be feasible. Because only one value is obtained for a given altitude, it is not possible to interpolate the distribution of a parameter for any value of α at this altitude. The method will systematically produce artifacts (isolated structures with vertical bands) unless the layer probed by the sUAS has a thickness at least greater than the vertical distance covered by the instrument to make 360°. It turns out that all the plots and discussions are incorrect, and section 3.4 should be deleted in its entirety, unless the authors can demonstrate the merits of their approach.

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 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.

Specific comments:

Line 8: see comment (1) above.

We have updated the abstract to be more precise

Line 34-41: Some important references on in-situ measurements of turbulence are missing, e.g., Barat and Bertin (JAS, 41, 819-827, 1984, and references therein), Bertin et al. (Radio Science, 32, 791-804, 1997), Alisse and Sidi (JFM, 402, 137-162, 2000), Gavrilov et al. (Ann. Geophys., 23, 2401–2413 2005). In addition, the manuscript ignores references from the radar literature (e.g. Sato and Woodman, JAS, 39, 2546-2552, Fukao et al., JGR, 1994, and many others). UHF or VHF clear air radars have enlightened the layering structure of the stratosphere, mentioned line 37, to be attributed to thin and horizontally extended turbulent layers or/and stable layers. Incidentally, the authors recognize that the horizontal stratification is a key feature of the stratosphere. Is this feature consistent with the ”vertical structures” supposed to have been detected with the sUAS measurements in section 3.4?

We have updated the literature review with these references

Line 65-67: The reference list about turbulence measurements from sUAS must be thoroughly revised. Some of them are not about turbulence (Bäfuss et al., 2018; Rautenberg et al., 2018, Jacob et al., 2018) and many others are missing. For example: Lawrence and Balsley (JTECH, 30, 2352-2366, 2013), Balsley et al., (BLM, 147, 165–178, 2013), Balsley et al. (JTECH, 35, 619-642, 2018), Reuder et al. (Acta Geophysica, 60, 1454-1473, 2012), Shelekhov et al., (Atmos. Ocean. Phys., 57, 533-545), Kanthe et al., (PEPS, 4:19, 2017), Luce et al., (JAS, 77, 231-2326, 2020), Calmer et al., (AMT, 11, 2583-2399, 2018), among others.

We have updated the text but note that Barfuss 2018, Rautenberg 2018 and Jacob 2018 do discuss turbulence measurements (although their focus may not have been on detailed analysis of the measured statistics) and Balsley 2013 was already cited. We have added most of the additional suggested citations.

Line 86-87: “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.” The comment is unclear. What do the authors mean?

Our intent was to point out that by advecting with the wind, statistics like horizontal spectra and structure functions become more difficult to calculate. We have revised the text accordingly.

Line 91-93: The introduction of the “infrasonic microphone” has already been made in the previous paragraph. It is this redundant. In general, the various paragraphs of the manuscript should be better organized to avoid such redundancies (they occur several times). This gives the impression of a juxtaposition of paragraphs with no guiding line.

This paragraph is introducing our specific experiments, and therefore we are specifically indicating that such a microphone was used in the present experiments. The previous paragraph was a general review of previous studies using infrasonic microphones. It is not clear to us what redundancy is being referred to here. We have made some adjustments to the text to avoid further confusion with the readers.

Line 121: Please convert km/h into m/s. The controllability of the sUAS is an important parameter and more information about limitations and performance should be given. We understand that the UAV can safely fly for wind conditions up to 31 m/s at least. How is the horizontal velocity of the glider controlled? It must be significantly high than 31 m/s for stability. A figure showing the ground speed of the sUAS with altitude (and the relative air speed, for the reason described in (2) Major comments) would be informative.

We have added airspeed/groundspeed figures to the revised manuscript and changed the km/h to m/s.

We have also added text describing how, the autopilot was set to maintain kinetic energy, and typically set near the optimal lift over drag ratio (the maximum distance that can be travelled per loss in altitude). To maintain the set airspeed the autopilot adjusts the pitch angle (the angle of attack of the mainwing airfoil). The horizontal velocity is a resultant of setting the airspeed and may fluctuate slightly from the pitch angle adjustments. Also, as the aircraft descends in altitude the air density increases and the HiDRON’s aerodynamic performance improves; thus, the horizontal velocity gradually decreased as the aircraft descended.

Line 129: please explain why the sampling is made at 10 Hz.

The HiDRON utilizes a radio modem for the command-and-control link and telemetry. The radio communication is also employed to send flight parameters to the ground station where the flight data is logged. Based on the modem model (Microhard P400), 10 Hz telemetry is selected for maximum efficiency to transmit and record data packets and to maintain reliability of the radio communication and command-and-control link.

The text was updated with this additional description.

Line 192: The authors seem to indicate that the five-hole probe sensor had an effective time response of 0.1 s. The reason is unclear (but I do not have the background to understand). The corresponding spectra should show a gap from ~ 10 Hz. It is roughly observed in Figure 8 at z=10 km (but around 5 Hz) and there is no evidence of a transition at z=1 and 18 km. How do the authors interpret this feature?

These sensors experience viscous damping in the tubing, as well as resonance in the transducer cavity which impact their frequency response characteristics. Therefore each probe has slightly different response characteristics and this particular sensor response was measured directly by introducing a step change in pressure and measuring the settling time of the tubing/transducer system by sampling the data at high rates (30kHz) during the step change.

We do not expect a gap in the spectra as the effect of the tubing is to damp out the response of the probe to fluctuations and this effect is countered by a resonance in the transducer cavity at higher frequencies (100 Hz in this case). The net combination tends to counteract each other and make it difficult to identify where in the spectrum inaccuracies are introduced by probe's response characteristics without measuring the response.

Line 227: The reason(s) of the choice of large values in circle radius (1-5 km) is not explained.

The 5 km radius was selected as a comprise between optimizing aerodynamic efficiency of the aircraft in a turn - by minimizing the bank angle and for safety to stay in proximity to the landing runway. For the 5 km radius turn the bank angle was approximately 5 to 7 degrees, and increasing the turn radius further would provide only a slight change in the bank angle. We have revised the manuscript to include this description.

Line 229: A figure showing the descent rates of the sUAS with altitude would be useful to figure out the conditions of sampling.

We have added this figure.

Line 243: "horizontal distances": with respect to the ground? If yes, it means the ground speed of the sUAS varied between 10 m/s and 80 m/s but this information is not provided in the manuscript. If these distances are expected to characterize the largest scales sampled during 30 sec by the instrument, it is not correct, because the relative speed should be considered (see (2) of Major comments).

As noted, we have updated the segmentation of the time series to be spatially equivalent to 1500 m using the relative air speed at the start of each segment. We have also added figures showing relative velocity and ground velocity as a function of altitude.

Line 252-271: These two paragraphs must be re-written. They are particularly confusing and not rigorous. For example:

- the first paragraph seems to describe general properties for the 3 flights but the second paragraph focuses on flights 2 and 3. So, we deduce that the description made in the first paragraph is for Flight 1. In addition, almost all the statements are disputable or not well- introduced. The agreement within 10% is unclear (please add information/figure that corroborates this result) and it is not true for ¡U¿ at all altitudes for example. But the paragraph focuses on temperature profiles. So, does this quantification only apply to temperature? But then why introduce the paragraph with "with the exception of RH". . . ?

We have completely revised these paragraphs as we have added additional nearby radiosonde profiles to better illustrate how the aircraft measurements compare to the range of values measured by the NWS radiosondes.

- The altitude of the top of the boundary layer is estimated to be "roughly lower than 3 km" in the first paragraph, but up to ~5 km in Flight 3. First, these altitudes (especially 5 km) are not realistic even for convective boundary layer. Second, the criteria used to estimate this altitude are not explained. Third, there is no indication consistent with these estimates in the figures.

These values are likely an oversight, being remnants from early drafts where we were using altitude in MSL rather than z, which we define as AGL which would result in a 1400 m difference. We had also only used the elevated TKE as a rough indicator of the PBL height. In the revised manuscript we have used values determined by more rigorous methods to estimate the PBL height and updated the text accordingly.

- The figures are not correctly labeled: "Fig. 5(a,c,e,)" should be "Fig 5 (a,b,c)". 'Fig. 5c" should be 'Fig 5b'. Fig.5 b, d, f should Fig 5 d.e.f. and at other places. Please check.

Thank you for catching this. Once again, this is due to a change made in the figure organization while drafting the manuscript but failed to update the corresponding text. We have corrected the figure referencing accordingly.

- What is the criterion used to define the tropopause altitude? It is found at 11 km (line 258) (presumably for Flight 1 in the first paragraph), quite consistent with figure 5a, but indicated to be at 12.5 km on line 278. It is found at 13 km for Flight 2 while the temperature inversion is actually observed at 11.5 km in figure 5b. It is stated that it is at 14 km for Flight 13, but an inversion can be found at 12 km. How are quantified the lapse rates in the troposphere and stratosphere and what is the interest to estimate such values (and tropopause altitudes) if they are not compared between the instruments? This comment also applies to humidity and velocity profiles, since the text does not describe the differences and similarities between the profiles, but rather their characteristics, which is another objective. There are too many caveats.

Again, these values are likely remnants from early drafts where we were using altitude in MSL rather than z, which we define as AGL which would result in the observed 1400 m difference (rounded to 1500 m). We have updated the text and, as noted above, including additional radiosonde measurements and try to utilize these figures more carefully.

From lines 290: This part should be separated from the previous ones because it is not about comparisons between radiosonde- and sUAS-derived profiles anymore. It seems to me important to show N^2 (squared BV frequency) and shear profiles before describing Ri profiles, since one of the purposes of the manuscript is to assess the performance of the sUAS measurements. In addition, low Ri values can have different causes, i.e. a strong shear and/or low N^2. The knowledge of these two parameters can help the interpretation of the turbulent events.

We have moved the Ri discussion to a new section and added corresponding profiles of N^2 and S^2 which allow better interpretation of the static conditions and wind shear present during each flight.

Equation (7): In practice, the impact of the variation of g with altitude can be ignored. The error is much less than all the other uncertainties.

We have reverted to $g = const$, in accordance with common practice.

Line 301: "Here we assume the critical Richardson number takes on a value somewhere in the range 0.25¡Ri¡1". It is unclear. Ric=0.25 is a necessary condition below which air can become dynamically unstable and turbulent (if Ric¿0.25, a shear instability cannot develop). Once turbulent, there is a critical Richardson number at which the flow begins to laminarize: it is generally accepted to be between 0.2 and 1 but turbulence can be found for $Ri >> 1$ according to Galperin et al. (2007) (but the corresponding turbulent regime should strongly differ from the turbulent regime for small Ri values). In practice, these thresholds must be used with caution because the Richardson numbers estimated from in- situ data are scale-dependent, i.e. depend on the vertical resolution at which they are calculated. It is common to apply arbitrary thresholds (i.e. "Ri is minimum and small 0.25-1")

This was our intent with specifying critical Richardson number as lying within a range and broadly defining this as a range of possible values below which turbulence could develop. Note that the revised Ri calculation approach has significantly increased the measured values above this critical range and therefore much of the discussion of critical Ri was edited out as it is no longer relevant.

Figure 8: the information is interesting. Why not showing the corresponding spectra for v and w? In Figure 8b and 8c, it should be F instead of Φ (y label).

The only reason v and w weren't included was simply to maintain clarity of the figure. However, in the revised manuscript we have decided to present the spectra in wavenumber domain, rather than the frequency domain, as it better fit with the discussion of inertial subrange slope at which they are introduced.

Line 333: I am again skeptical about the interpretation of the increased TKE layer up to 4 km as corresponding to the CBL top

As noted above we have applied a more rigorous definition of the boundary layer height in the revised manuscript, we have also made sure to bette distinguish being near the boundary layer to being within the boundary layer.

Figure 9: EDR and TKE should be presented in logarithm scale (and continuous lines), because the linear scale over-represents the maxima near the ground. Indicating that "TKE is close to 0", line 333, is symptomatic of the fact that the linear scale is unsuitable for the present purpose. Figures showing the slopes, as calculated in Figure 8 for 3 cases, vs altitude and for the 3 wind components should be included. It would enable us to identify the altitudes where the inertial slope is indeed observed, those where a different regime is observed, and those where instrumental noise is dominant for all frequencies. This information is essential for the purpose of the manuscript. A characterization of the slopes vs other parameters (e.g. TKE, EDR, Ri, etc) could be very informative.

As noted above, we have changed our presentation of all figures to use lines instead of data points. We have also changed the profiles of TKE, EDR, and Ri to use logarithmic axes. Finally, we present the distribution of measured slope in Section 3.5.

The dataset offers the possibility to show $\langle u'^2 \rangle$, $\langle v'^2 \rangle$ and $\langle w'^2 \rangle$ separately. Plotting, for example, $\langle u'^2 \rangle$ vs $\langle w'^2 \rangle$ would be interesting for quantifying anisotropy. A discussion of this anisotropy in light of ε, Ri, etc, would be very enlightening.

TKE and ε are related by a master length scale (e.g. Mellor and Yamada, Rev. Geophys. Space Phys., 20, 851-8751982) which of great interest for the characterization of turbulence. The dataset shown in Figure 9 offers the potential to estimate this scale.

We agree with both these sentiments, and following submission of this manuscript have continued examining these and other aspects of the data. However as the current draft of the paper is already over 30 pages, we feel that an in depth statistical analysis of the data set is beyond the scope of this initial paper (intended to describe the measurement system, measurement approach and capabilities.)

Line 337: "...caused by inertial turbulence" and remove "elevated" in the same sentence because the comment is valid for all levels of TKE. The spectra with a -1 slope may either reveal another turbulent regime or be due to a white noise contamination even for $f < 5Hz$ when the atmospheric signal is weak. Figure 8c is apparently in favor of the first interpretation for the selected case but it is not necessarily always true, especially when the instrumental noise dominates.

Our expectation is that the instrumentation noise would be indicated by a noise floor on the spectra which does not appear until much higher frequencies than those presented. As the measured spectra are consistently above this noise floor (with content over an order of magnitude higher), we do not expect this deviation to be due to white noise contamination (at least from instrumentation noise) and instead expect it to be due to a different turbulent regime.

Line 343: "... more active turbulence conditions during these flights". This statement should be nuanced because (1) the rejection was based on the u spectra only, (2) the non- detection of a -5/3 slope does not mean the absence of turbulence, (3) the corresponding levels of EDR of flight 1 (qualified as "weakly active") shows a significantly higher background than flight 2, indicating higher spectral levels but not consistent with an inertial subrange. As we do not know about the interpretation of the observed non-inertial subranges, turbulence activity cannot be qualified.

This was only meant to broadly refer to the increased scatter in the TKE profiles. We have revised the text to be more clear.

Line 345: Please remove this sentence. The reference of Kelvin waves is not suited here because equatorial Kelvin waves (Fujiwara et al. 2003) are waves trapped around the Equator similarly to coastally-trapped Kelvin waves.

We have revised as suggested

Equation (9): The notation Φ_{11} of the spectrum may not be appropriate ($F_{11}(k_1)$ would fit better the notation used in Figure 8). Φ_{ij} generally refers to the spectral density tensor (see e.g. Doviak and Zrnic', Doppler radar and weather observations, p. 326, 1984). The sUAS is "sensitive" to the 1-D longitudinal spectrum (see e.g. Hocking (EPS, 1999)).

We intentionally defined F to indicate velocity spectra defined in the frequency domain and Φ to indicate velocity spectra in the wavenumber domain (different notation is required as $\Phi \neq F$ due to the conversion required to ensure variance is preserved when both integrating F in f and when integrating Φ in κ). Note that the turbulent velocity spectrum tensor is also commonly referred to as Φ_{ij} in the turbulence literature (see textbooks by Pope or Tennekes and Lumley, for example) which is why we used it here. However, to avoid conflicting with established nomenclature, we have replaced usage of $\Phi(\kappa)$ with $E(\kappa)$.

Line 367 refers to $\Phi_{11}(k_1)$ (Φ_{11}?) and F (F_{11}?). Line 368: k_1^n should be $k_1^{-5/3}$. Strictly, the power-law fitting should be applied to a limited range of k_1 since the smallest wavenumbers are not well-resolved.

Line 367 was describing the conversion described above and therefore required both $\Phi_{11}(\kappa_1)$ and $F_{11}(f)$. Line 368 wass specifically describing the power law fit (which does not pre-assume $n = -5/3$, as we are using it to evaluate how closely the inertial subrange slope, n is to -5/3, and we also specifically note that the wavenumber range is limited to that where the probe response is reliable. For the 30 s average we found that a lower frequency bound was not necessary for the fit, however the larger statistical windows used in the revised calculations required implementation of a low wavenumber bound of $\kappa_\ell > 0.1$ m^{-1}.

Technical comments:

Line 4-5: The sentence is unclear, please rephrase.

This sentence has been revised.

Line 20-21: please add references.

Citations added.

Line 24: "Despite the higher stability of the stratosphere"

This sentence has been revised.

Line 25, see also line 318: "... due to mechanical and thermal disturbances" -¿ "due to shear instabilities and gravity wave breaking". The formulation is unsuitable because the mechanical sources of turbulence refer to those produced by obstacles close to the ground. The rest of the paragraph is awkward –and not rigorous- and references of the "classical" literature should be included instead.

This sentence has been revised.

Line 38: “high turbulent kinetic energy dissipation rate”: please be more quantitative.

Revised with the addition of a quantity.

Line 45: “and for identifying the inner scale of turbulence.”

Our preference here is to retain finer scales (as referencing the fine scale structure of turbulence, as used by Townsend for example) as opposed to ‘inner scale’ since inner scales have very precise definitions in boundary layer turbulence which are not appropriate here). We also considered using the term microscale structure, consistent with how Kolmogorov scales are typically described, but were concerned that this may conflict with the ‘microscale’ spatial and temporal scales used in meteorology.

Line 45: “This experiment”: please be more specific with references.

This sentence has been revised.

Line 45-47: The Richardson number is not defined and it is not explained why Ri=0.25 is an important value and why turbulence when Ri ¿ 0.25 should be noted. The authors should indicate that some LITOS results were corrupted (Soder et al., AMT, 2019) due to balloon wake and that turbulence observed when Ri ¿¿ 1 was suspect. In addition, useful information on the relationship between Ri and ε can be found in earlier references mentioned in the specific comments (line 34-41).

We have moved the mathematical definitions of Ri, ε, and N^2 into the introduction and added some text introducing the critical Richardson number.

We have als added the Soder et al reference to the discussion of the LITOS experiment.

Line 63-72. Pitot tubes are also used (e.g. Lawrence and Balsley (JTECH, 30, 2352-2366, 2013).

This sentence has been revised to include Pitot tubes.

Line 90: “air masses” generally refer to “large bodies of air” at synoptic scales in meteorology. The term is not suitable here.

Replaced with ‘turbulent eddies’ which more aptly describe our intended meaning.

Line 90: “geostationary” usually refers to satellite orbits. Do the authors mean “relatively constant location above the ground?”

That is indeed what we meant, we have revised the text as suggested

Line 92: “”traditional” → “standard”

Revised accordingly

Paragraph 2.1: It seems more natural to present the instruments first, then the configuration of the experiment. 2.1 → after 2.3.4 and before 2.4.

Although we felt that 2.1 served as an introduction to the overall experiment, we have moved the section as requested.

Fig.1 : Please add the location of the balloon launch site (El Paso) and show the distance in km (in Fig. 4 also). The distances are crucial for the interpretation of the radiosonde and sUAS data and longitude/latitude

coordinate system is of little use here.

The balloon launch site was well outside both these figures and would not be suitable to include. Instead, we have added an Appendix with upper air wind maps and satellite imagery on which the radiosonde launch sites were indicated

Line 102: “Three sUAVs were flown” (?)

Three flights were conducted with the same sUAS. We have updated this sentence to be more clear.

Line 106: The altitudes are given in km m.s.l but the profiles are shown from z=0 (i.e. above the ground (line 241). Please indicate the corresponding altitude AGL, even if it can be roughly deduced from Fig. 1.

We have added the launch/recovery altitude to the experiment description and added the a.g.l. release altitudes to Line 106.

The third sUAS was released from 30 km m.s.l (28.5 km AGL ?), but the profiles are shown up to 25 km. Please clarify.

Although the release was at 30 km, it took about 3 km of altitude before the aircraft returned to its controlled orbit (See figure 4c). For consistency, we only report data only from the controlled orbital descent phase of the flight.

Line 191: “. . . the actual probe frequency response. . . ”

Revised

Line 194: remove “disconnected”

Revised

Lines 232-235: Please indicate Local Time instead UTC (and avoid MDT). By doing so, the reader does not need to convert by himself when interpreting the PTU profiles measured by the sUAS and the radiosondes at different times (Figure 5).

We have provided launch/release/recovery times in UTC and LT in the revised manuscript.

Line 246-249: The first sentence is not necessary and the second has already been written. A more detailed description of the radiosonde data is necessary (see (1) of major comments).

We have revised the text accordingly.

Figure 5: please add LT times for all the radiosonde and sUAS flights. A figure showing the trajectories of (and horizontal distance between) both instruments is necessary.

We have added the LT as requested, however a figure with trajectories is not feasible due to the distances involved (glider orbits were only 10km diameter). Instead when introducing the radiosonde data, we clearly indicate the separation distances and involved

Line 258: For flight 1, “the temperature continued to decrease with altitude at a rate of 1C/km” → “the temperature continued to decrease at a mean rate of 1C/km between 11 and 19 km” (otherwise it is confusing, see specific comments also)

Revised.

Line 274: please add “(not shown)”.

Revised.

Line 281: please show the NOAA upper air wind maps. The absence of reference points make difficult to confirm the statements.

We have added the NOAA upper air wind maps to Appendix A.

Line 282-285: what do the authors mean ?

We simply were explaining why the jet stream was no longer evident during Flights 2 and 3. We have updated the text to improve clarity.

Line 307: The term “potential instability” refers to “an atmospheric condition in which otherwise stable air would become unstable if forced to rise (e.g. over high ground) thereby reaching its saturation point.” See e.g. www.encyclopedia.com. Please replace “potential” by “possible shear”.

This sentence has been revised. We had only meant to indicate that there was potential for an instability to develop.

Line 309: “marginally unstable tropopause”: what do the authors mean?

We were referring to the lapse rate classifications for the troposphere. We inadvertently replaced troposphere with tropopause. This typo has been corrected and we have added a reference for the lapse rate classifications.

Line 316: Here again, the terminology is improperly used. “An atmosphere is said to be “conditionally unstable” if the environmental lapse rate is between the moist and dry adiabatic lapse rates. This means that the buoyancy (the ability of an air parcel to rise) of an air parcel depends on whether or not it is saturated.” (see glossary of meteorology) . “suggesting the possibility of localized buoyant production”. Do the authors refer to statically unstable conditions, ie. Ri ¡0? If yes, it must be clearly stated and defined earlier, e.g. around line 302.

This sentence was originally written in a slightly confusing manner. The conditions for buoyant production in the boundary layer were referring to the observations $Ri < 0$ in the boundary layer. The reference to the ’conditionally unstable troposphere’ was referring specifically to the measured lapse rate, although it was not clear in the sentence as written. The revised Ri and N^2 calculation approach reflects statically stable conditions were present throughout the altitudes measured (except the boundary layer) and therefore this sentence was no longer relevant.

Line 318: “mechanical turbulence”: see above, comment for line 25

This sentence has been revised.

Line 328: Do the Hanning window preserve variance?

Yes. This was confirmed prior to use.

Line 329: Do the cut-off at 5 Hz related to the effective limited time response indicated line 192?

Yes. We cannot have confidence in the frequency content measured above 20 Hz and the initial 5 Hz threshold was selected to provide additional confidence in the range used. However, to increase the data points included to the calculation we have revised the integration range to include frequency content up to 20 Hz.

Line 334-335:The description is unclear because of the use of a linear scale and a dot representation (see “specific comments”)

We have updated all the figures to use lines and logarithmic axes where appropriate.

Line 336: Please remove "although the regions of elevated k appear at different altitudes". This comment is not useful.

Revised.

Line 344: Please indicate the altitude of the tropopause in Figure 9.

The logarithmic presentation of the updated figures obscures the previous observation of an increase in k and EDR near the tropopause, therefore we have removed this statement and feel adding an indicator of the tropopause location to the figure unnecessary.

Line 354: "As direct measurements of . . . " please explain more or add a reference

We have revised the text to expand our explanation

Line 355-356: The sentence indicates a condition that has no reason to exist at this stage, since inertial domains have been identified by spectral analysis.

Please note that we are applying this calculation to all 1500 m long segments within the time series to determine ε, hence we need to formally assume equation 9 is at least approximately valid in order to obtain an estimate of ε for each segment, even if no inertial subrange is evident.

Line 520: Please remove "flux". The gradient Richardson number and the flux. Richardson number have two distinct definitions.

We have corrected this typo and are aware of the distinction.

Response to Referee 3

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-Väisälä 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".

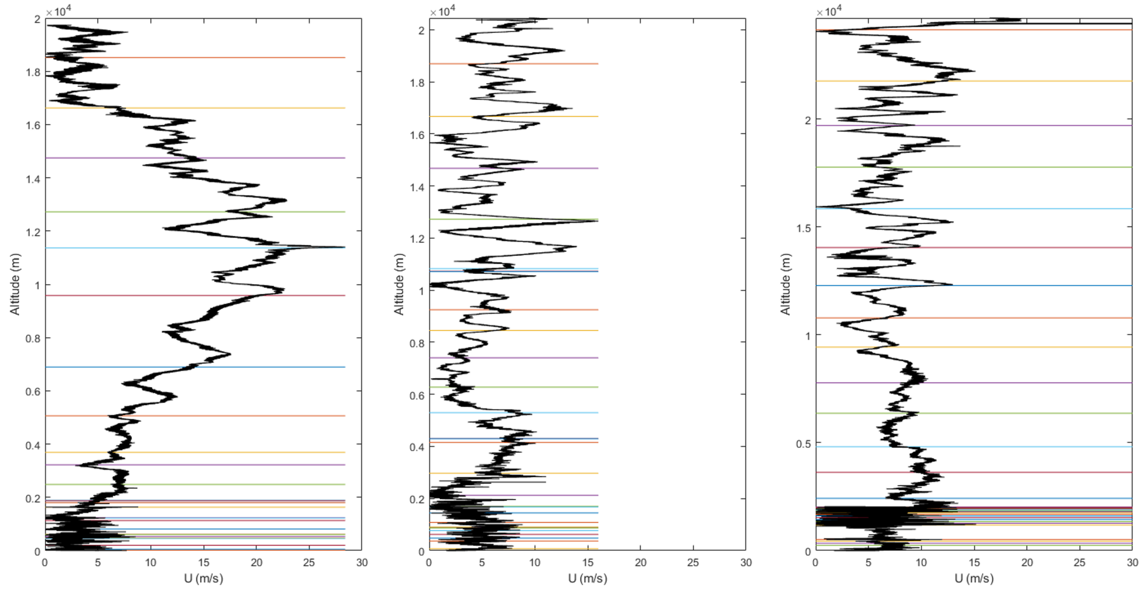

Figure 1: Vertical profiles of wind magnitude from all three flights shown with horizontal lines indicating location where aircraft heading passed through 180°. (left) Flight 1 (middle) Flight 2 (right) Flight 3

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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-Väisälä 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.