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

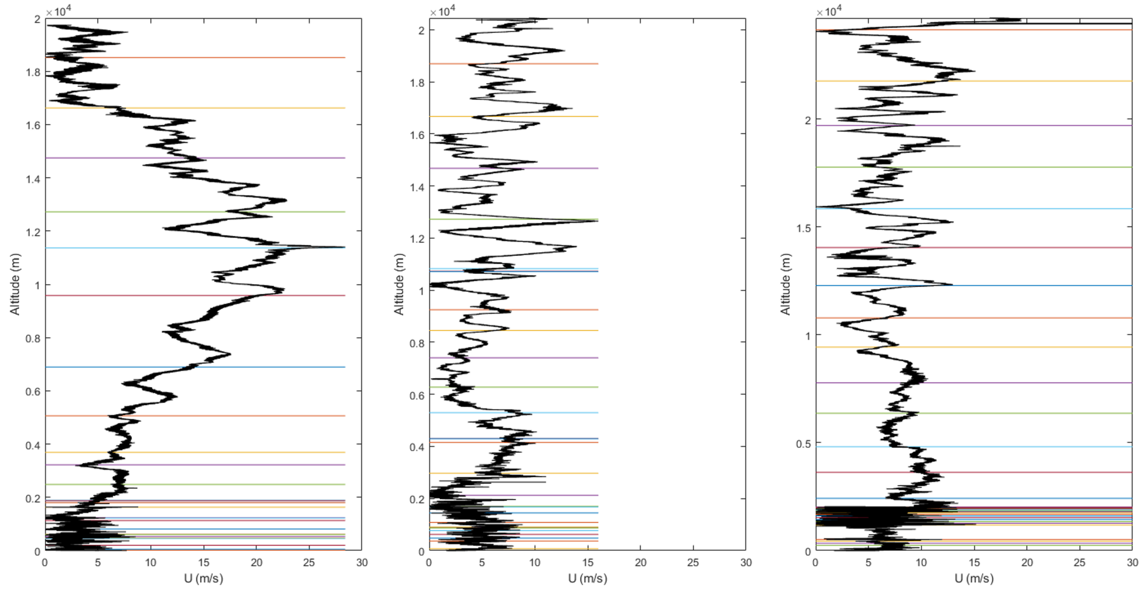

Figure 1: Vertical profiles of wind magnitude from all three flights shown with horizontal lines indicating location where aircraft heading passed through 180°. (left) Flight 1 (middle) Flight 2 (right) Flight 3

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 $d\lambda = Rd\alpha$ 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.