We thank the referee for its constructive remarks which should undoubtedly contribute to improve the present paper.

Follows a point-by-point response to the referee's remarks.

Major remarks

1. The sampling period of measurements is 30 s, i.e. $f_{Nyquist}=1/60$ Hz. The oscillation periods of the balloon vary around 220 s, i.e. $f_{NBO}\approx f_{Nyquist}/3.5$. We found that the variability of measured quantities in this narrow frequency domain is dramatically affected by balloon oscillations (see the *w* spectrum of Fig.6). This conclusion is given in lines 200-204 and 401-404. In fact, our first attempt to detect turbulence was to look for an excess of variance or a spectral signature associated with fluctuations of the measured parameters in the $f_{NBO}-F_{Nyquist}$ frequency range. This method was unsuccessful. The main reason, we think, is that the high-frequency fluctuations are dominated by the balloon's natural oscillations. Therefore, we turned to statistical methods to detect the effects of turbulence from sensors that do not allow to directly measure turbulence.

2. As noted by the reviewer, the purpose of this paper is not to describe geophysical results. We indicated however that the detection is not uniform around the globe but seems to depend on the position. The following figure showing the occurrence frequency of the turbulence detection supports this statement. The variability according to the position appears to be very large, ranging from zero to about 25%. However, we think that the interpretation of these results is beyond the scope of the article and we do not think to publish it here. A PhD thesis by one of the co-authors (CP) is in progress on the geophysical exploitation of strateole-2 data.

We furthermore think that the turbulence we observe in the stratosphere is primarily due to wave breaking. As commonly accepted, deep convection is a major source of waves in the tropics, and those waves propagate vertically in the atmosphere. We therefore do not expect much difference between the occurrences of turbulence in the TTL and STR balloon flights: those are only 2 km apart. On the other hand, we expect a significant longitudinal modulation.

From Table 5 indicating the fraction of time during which the flow is turbulent does not show any clear contrast between the TTL and STR balloons, but it is not conclusive as only two TTL balloons are reliable (excluding flight 3, very noisy).

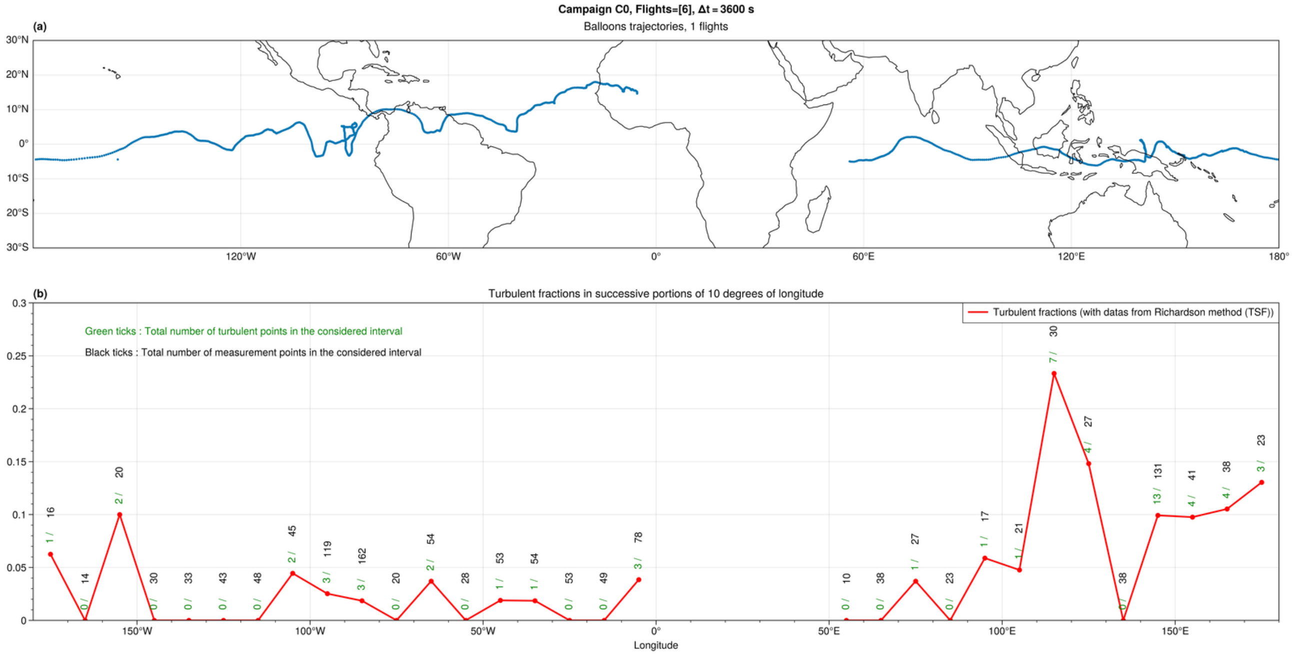

Figure a: Trajectory of flight 6 (STR1) of the C0 campaign (top) and percentage of turbulent detection by the Ri method (bottom).

3. We thank the reviewer for pointing out the (too) many typos for which we apologize. The mentioned typos – and others - have been corrected.

Minor remarks

- **line 10**: The mode of the distribution of the vertical displacements associated with the oscillatory motions of the SPBs is ~ 15 m (see figure 4).

- **line 30**: Gravity waves transport vertically energy and momentum. As long as waves do not saturate, there is no impact on the dynamics or chemistry of the atmosphere. Heat, momentum and minor-constituent fluxes occur when wave saturation occurs, i.e. when waves break into turbulence.

- **line 75**: We wrote "under high pressure balloons because the measurements are carried out on board of a gondola and temperature sensors located under the balloons.

- **line 115**: we downgrade the RACHuTS measurement to 30 m 1) to improve the measurement precision, and 2) in order to reach similar vertical resolution for TSEN and RACHuTS, TSEN estimates being evaluated in layers of about 30 m thick (+/- 15 m) -.

- **line 125**: The instrumental noise level is quantified by the standard deviation of the high frequency fluctuations. This standard deviation is estimated from the variance of the nth order increments as described in Appendix 1. Because both the balloon motions and the instrumental noise contribute to the high-frequency variability, we consider only the smallest 10% of estimates to minimize the influence of the balloon motions.

- **Eq 1**: $\pi \approx 3.14159...$

- **line 195**: The referee's remark is quite right. Most large amplitude oscillations are associated with depressurization events. We have added a sentence to clarify this fact:"A few large amplitude oscillations (>100 m) are observed. We found that they are most often associated with depressurization events. ".

We did not handle balloon depressurization events because they do not appear to have impact on turbulence detection (we do not detect turbulent events during depressurization events).

- **line 220**: Since the horizontal wind is typically of the order of 5 m/s, the horizontal spatial scale is typically 18 km. But this horizontal scale does not correspond to the spatial extension of the observed episodes (turbulent or laminar) since the balloon is advected by the flow, i.e. the observations are Lagrangian. Lagrangian observations of turbulent or laminar events are rather related to the lifetime of the events.

- **Fig 5**: The red line at t_1 indicates the potential temperature level that the balloon will reach at t_1+30 s. It is drawn to indicate schematically that the whole layer is vertically displaced in the 30 s time interval. Therefore, the observed θ changes are not simply proportional to the vertical θ gradient.

- **Fig 6**: The two spectra in Fig. 6 are representative of the power spectra of *w* for all flights: almost flat up to the frequencies corresponding to the oscillatory motions of the balloons. Similar power spectra have already been published (Podglagen et al.).

- **line 270**: Thanks to the reviewer for pointing out the (too) many typos. We have corrected those mentioned, and others.

Fig 7: The colors of lines of the plot have been corrected

line 332: “the time series” → ‘parameters’

line 374: The probabilities are indicated as the y-axis of the cumulative distribution function (bottom right). The percentages are deduced.

line 387: The differences about the turbulence detections from the different methods are discussed in the paragraph “Discussion”, section 5.1 and are shown in the new figure 14.

The detection methods are based on statistics (two correlations and two regressions) applied on independent measurements. The percentages of turbulence differ by using different estimators or different parameters: the percentage differences can reach a factor of two (table 5, flight 07_STR2, 6.3% vs. 3.3%). These differences can be partly due to the thresholds values of the hypothesis tests of a null correlation (choice of a confidence interval). Likely more important is the fact that the thresholds values (null correlations or $Ri=0.25$) are associated with the tails of the distributions of these statistics (see the histograms of figures 9 and 13). The differences occur mainly when the atmosphere is close to a neutral stratification, i.e. when the estimates are close to the thresholds. When the atmosphere is somewhat stratified (about 95% of the time) the statistical estimators are very consistent, the agreement between the estimates being larger than 99% (Fig. 14 of the paper).

The following figure shows the turbulent periods detected by the four different methods for flight two. As stated in the paper, we privileged the combination of T_C and z_P as they

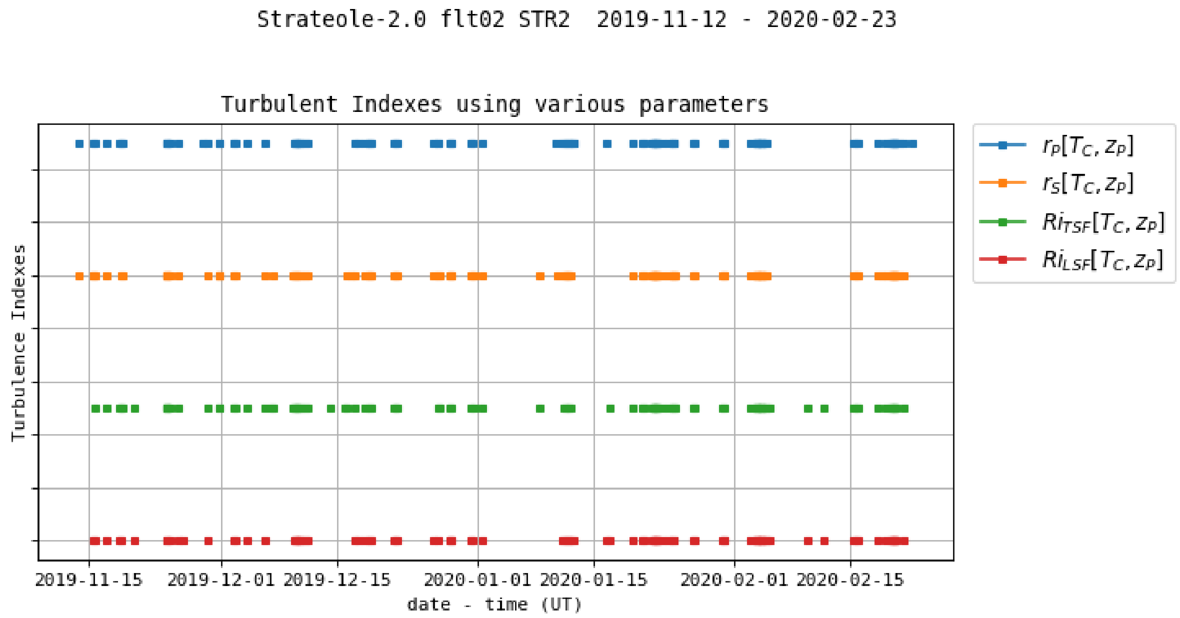

Figure b: Time series of the turbulence detection from four estimators.

appear to be less noisy than T_S and z_{GPS} . Overall, the agreement is good, even differences are visible.

In short, the used statistics (correlations and Ri) allow to describe the state of stratification of the atmosphere in which the balloons float. When the atmosphere is close to a neutral state, all estimates, correlation or *Ri*, reach values belonging to the tail of their distributions, close to threshold values. Some differences in detecting the turbulent episodes occur due to threshold effects. We added a subsection about the comparison of the different indexes (.

Appendix A: The two typos in the equations have been corrected.