1) One major concern about the analysis: Lines 62-75: As far as I can tell, the definition of the segments of length "s" implies that not the same data is used for calculating F(s) for each "s" as "N" is not a multiple for all "s". How does this affect the results, in particular, for larger s (for instance, s> N/2)?

Response： As the referee commented, the amount of data used for calculating $F(s)$ for each 's' is different. This is a common problem that many time-series analysis methods encounter. Generally, as long as the scale 's' is not particularly large, it is commonly considered that results would not be seriously affected by the amount of data. The same is true for the DFA. According to numerical simulations of many mathematical models, Kantelhardt et al. (2002) recommend that $s < N/4$, because for large 's' statistical errors become large. In practice, one can approximately define the maximum of 's' for fitting by the position where $F(s)$ fluctuates significantly around the power law. From Fig. 2 in our paper, one can see that the large fluctuations of $F(s)$ appear at $s \approx 1000$ sec. Thus, the maximum of 's' for fitting is set to be 1000 sec in our analysis. Because the crossover scale is about 20-30 sec that is far less than the maximum of 's' for fitting, the statistical errors caused by large 's' would not affect the results in this study.

In order to facilitate readers to understand the analysis method of this paper, above discussions have been briefly added in lines 75-76 in the revised paper.

Kantelhardt JW, Zschiegner SA, Koscielny-Bunde E et al. (2002), Multifractal detrended fluctuation analysis of nonstationary time series. Physica A, 316: 87–114.

2) Another concern is that, although being a reasonable approximation to the observations, the simple OU process could be fit better through some extensions (as mentioned in lines 144-145). Why are they not tested in this study? At the very least the physical reasons for the difference between the observations and the OU should be discussed? These physical insights should inform the extensions the OU requires in order to fit the data better.

Response: There may be some misunderstandings about our statements in lines 144-145. We meant to say that the OU process can be used as a starting point for future study. Based on our analysis of the OU process, one would find a certain kind of extension to fit the data better than the OU process. However, until now we still don't know this kind of extension. To avoid misunderstanding, we have deleted these statements in the revised paper.

Our paper indicates that the intrinsic non-stationarity (IN) is intimately related to the inertial-subrange turbulence which is normally considered to be generated by the cascade process. After all, the OU process is just a simple mathematical model and does not consider the cascade process. Thus, we also believe that the extension could be found by adding the cascade process into the OU process. This paper only focusses on the main characteristics of IN. The complicated physical mechanism of IN and the corresponding improved fittings would be reported in future work.

Above discussions have been briefly added in lines 160-164 in the revised paper.

3) Line 7: "... we find that the average time should be..." - Do you mean the average time for eddy covariance measurements?

Response: Yes, it is the averaging time for eddy covariance measurements. To be precise, it is the averaging time used to average the instantaneous turbulent flux and is usually called the flux-averaging time in literature. We have clearly interpreted this term in Sec. 3.3. For clarity, we have revised 'the average time' to 'the flux-averaging time'.

4) Lines 11-18: The first paragraph and the first two lines of the second paragraph are not related to the rest of the paper. The paper deals with quantifying the intrinsic non-stationarity of carbon fluxes. Such quantification is not related to climate change nor does a better understanding of it allow for reductions in anthropogenic greenhouse

gas emissions. In particular, lines 17-18 are unclear how a better understanding of the CO2 transport is helpful in reducing the CO2 reductions. I guess the authors refer to air quality and not reduction in CO2 emissions.

Response: As the reviewer commented, the first paragraph is less related to the rest of the paper. Thus, we decided to delete the first paragraph in the revised paper.

The first two lines of second paragraph are intended to briefly explain the significance of the research subject of this paper, that is, the CO2 vertical exchange. However, the sentences are too ambiguous to cause misunderstanding. We have rewritten these sentences in the revised paper.

5) Lines 93-94: "The time series seriously contaminated by high-frequency white noise are also removed?" Why? How does white noise affect the analysis?

Response: We remove the high-frequency white noises because they are not real signals of wind speeds and carbon dioxide concentrations. At high frequency, both the spectra of wind speeds and carbon dioxide concentrations should obey a power law with the power exponent of -5/3. However, the spectrum of time series seriously contaminated by high-frequency white noises will be flat at high-frequency range. The white noises may be caused by the random electrical signal noise of anemometer or the mechanical resonance or other reasons. In this paper, we do not pay special attention to the effect of noise and any other kind of problematic data. We just discard all of them in the analysis, as usually done in calculations of turbulent fluxes.

6) Line 103: Why do you choose these Reynolds averaging timescales? In particular, 6 seconds is not a standard value to decompose the turbulent fluxes from the mean flow.

Line 110: "The crossover scale in the case with 6 seconds is smaller than that in cases with 900 and 300 seconds." - what does this tell us? What is the physical interpretation?

Lines 114-115: "Results indicate that the IN is a small-scale phenomenon which is intimately related to the inertial sub-range turbulence" - Isn't that expected?

Response: The Reynolds averaging timescales of 900s (15 min) and 300s (5min) are commonly used in the eddy-covariance method (for example, see Doran, 2004; Metzger et al, 2007; Li and Bou-Zeid, 2011; Donateo et al, 2017). The Reynold averaging filters out scales greater than the averaging timescales. The intrinsic non-stationarity (IN) is related to scales (crossover scale) much smaller than the averaging timescales. Thus, the fluctuation functions do not significantly change with large Reynold averaging timescales, as shown in Fig 3.

We note that the crossover scale is about 20-30 sec. Thus, if we use a Reynolds averaging timescale much smaller than the crossover scale, the crossover phenomenon would be changed. In order to show this phenomenon in sharp contrast, we choose a timescale of 6s far away from the crossover timescale. We indeed observed that the crossover moves to a smaller scale. Someone believes that if the sampling frequency is improved and the flux-averaging timescale is further reduced, the stationarity assumption of the eddy-covariance method can be better guaranteed. Our findings indicate that the above consideration may not be right because the further reduction of flux-averaging timescale would face another new kind of non-stationarity (that is IN) and would partly remove the contribution of turbulence.

Above discussion have been briefly added in the end of Sec. 3.1 in the revised paper.

The statement of L114-115 is indeed an expected conclusion. However, we still need to do spectral analysis to confirm this conclusion (see Fig. 3c). The significance of this conclusion is to tell us that the IN is not rooted in large-scale meteorological processes, but rooted in the nonlinear and non-Gaussian cascade process of inertial-subrange turbulence. It is the inherent property of turbulence. As we have said in the second respond, this conclusion gives us a possible direction to improve the OU process.

Doran, J.C. (2004). Characteristics of Intermittent Turbulent Temperature Fluxes in Stable Conditions. Boundary-Layer Meteorology 112, 241–255.

Metzger M, McKeon B.J and Holmes H (2007) The near-neutral atmospheric surface layer: turbulence and non-stationarity. Phil. Trans. R. Soc. A.365, 859–876.

Li, D., Bou-Zeid, E. (2011). Coherent Structures and the Dissimilarity of Turbulent Transport of Momentum and Scalars in the Unstable Atmospheric Surface Layer. Boundary-Layer Meteorol 140, 243–262.

Donateo, A., Cava, D. and Contini, D (2017). A Case Study of the Performance of Different Detrending Methods in Turbulent-Flux Estimation. Boundary-Layer Meteorol 164, 19–37.

7) Lines133-134: "..., although the fluctuation exponent of data seems to be greater at large scales and less at small scales, compared to the OU process." What does this tell us? What are the physical reasons for that behaviour?

Respond: As we have stated in the end of Sec. 2.2, if the fluctuation exponent $0.5 < \alpha < 1$, the time series is stationary and long-term correlated. If the fluctuation exponent $\alpha = 0.5$, the time series is stationary and independent.

The fluctuation exponent of the OU process at large scales equals to 0.5. The fluctuation exponent of data seems to be greater than that of the OU process but less than 1 at large scales, which indicates that the data is stationary and long-term correlated at large scales. This could be related to the large-scale coherent structure of scalar turbulence (Celani and Seminara, 2005; Liu and Hu, 2020).

At small scales, the OU process approaches the Brownian motion, that can be easily inferred from Eq. 8 in the paper:

$$\lim_{\Delta t \to 0} \Delta y = b\sqrt{\Delta t}\xi,$$

where ξ is an independent random variable with the normal distribution. Thus, the fluctuation exponent of the OU process at small scales equals to that of the Brownian motion. The fluctuation exponent of data seems to be less than that of the

OU process at small scales, which indicate that the data deviate from the Brownian motion at small scales. This could be related to the ubiquitous non-Gaussian intermittency of turbulence at the inertial subrange (Liu et al., 2019).

Above discussion have been added in the end of Sec. 3.2 in the revised paper.

Liu L., Hu F., Huang SX (2019). A Multifractal Random-walk Description of Atmospheric Turbulence: Small-scale Multiscaling, Long-tail distribution, and Intermittency. Bound.-Layer Meteor., 172, 351–370,

Liu L., Hu F. (2020). Finescale Clusterization Intermittency of Turbulence in the Atmospheric Boundary Layer, Journal of the Atmospheric Sciences, 77: 2375-2392.

Celani A. and Seminara A. (2005). Large-scale Structure of Passive Scalar. Phys. Rev. Lett., 94:214503

8) Line 4: "widespread" sounds as if a spatial analysis has been done which is not the case. I suggest using "common".

Response: We have replaced the word 'widespread' by 'common' in the revised paper. Thank you.

9) Correct the citation style. At most places the references should be (Author, year), e.g. line 15

Response: That is caused by a typesetting software error. We have checked the entire document carefully and corrected all errors in the revised paper. Thank you.

10) Lines 63: "m is a positive integer" - I guess 2<= m <= N? This goes back to general comment about the accuracy of estimating F(s) as s gets large

Response: Yes, m is a positive integer and 2<= m <= N. In practice, we choose a value of m much smaller than N to fit the $F(s)$, because the statistical errors get large at large values of m. The maximum of m for fitting is changed case by case.

We normally plot $F(s)$ with m from 2 to N. The maximum of m can be defined by the position where $F(s)$ begins to fluctuate around the power law significantly.

11) Line 84: "The data were…" suggestion: "Carbon dioxide turbulent fluxes were …"

Response: The 'data' are referred to the raw data including wind velocity and carbon dioxide concentrations. Due to this reason, we intend to use 'data'.

12) Lines 91-93: Hard to understand the 2 sentence: "The quality control methods, proposed by Vickers and Mahrt and including spikes…" Do you mean: "The quality control methods proposed by Vickers and Mahrt (1997) are applied to remove problematic data with spikes, dropouts, ..."?

Response: We have revised the statement as the referee suggested.

13) Figure 3: coordinate the colors between the different subplots such that 6 sec, 300 sec, or 900 sec have the same color in each subplots

Response: We have coordinated the colors between the different subplots in the revised paper.

14) Line 115: "The choice of very small Reynolds …" - word missing (averaging timescales?)

Response: Yes, some words are missed here. We have corrected this mistake in the revised paper.