Reaction to referee #2:

We thank the referee for his/her comments that helped improving the manuscript. We agree with his/her general comment that SICAS measurement results can and should be improved for future research. This is something that we have been working on in the last year and we expect to gather measurements of higher quality in the near future. Despite the low $\Delta'(^{17}O)$ variability in the atmosphere and the relatively high uncertainty of our measurements published here, the interannual changes in the $\Delta'(^{17}O)$ records of Lutjewad and Mace Head are significant, and should be studied further in order to improve our understanding of the $\Delta'(^{17}O)$ budget in atmospheric $CO_2$. We think the records, together with the comparison of the measurements and the model simulation we present in the manuscript contain valuable information on the potential influence of the stratospheric input of $\Delta'(^{17}O)$ on the total budget of $\Delta'(^{17}O)$ of $CO_2$ in the troposphere. Below we react on the referee's comments in the bold and italic text.

Major comments:

1. Need to have a paragraph summarizing the errors/biases of SICAS and possible sampling/storage biases. The SICAS D17O measurements/results are suspicious. Detail analysis of IRMS D17O, though limited, is not available. See specific comments below.

    ***We understand that the reviewer is skeptical, but we do not agree with the qualification that our results are suspicious, and that limited details of this technique are available. To help the discussion on this forward, we reiterate some of it here and also added extra text to the revised manuscript to help the reader appreciate the technique, and its uncertainties, better.***

    ***For more elaborate technical details on the SICAS errors/biases and possible sampling/storage biases we refer to Steur et al. (2021), Steur et al. (2023) and Steur (2023), but similar measurement and calibration techniques can also be found in Bajnai et al. (2023), Hare et al. (2022) and Perdue et al. (2022). In the manuscript we only discuss these matters briefly, as our main aim is to present the measurement records, not to describe the technical details of the measurement process.***

    ***The main reason for higher uncertainties of the SICAS measurements is due to $CO_2$ amount fraction dependencies, also identified in Bajnai et al. (2023), which are hard to correct when the samples are outside the range of our reference cylinders. We added the following text in line 140 that elaborates on this:***

*"We calibrate with the reference cylinders only, instead of having an on-line mixing facility where the reference and sample $CO_2$ amount fraction can be matched (Perdue et al., 2022; Bajnai et al., 2023). Therefore, samples that fall outside the range of the $CO_2$ amount fraction that is covered by the reference cylinders will have higher uncertainties."*

*We thereby added the following outlook on how to possibly deal with this in the future (line 171): "Extending the $CO_2$ amount fraction range of our reference cylinders will improve the measurement precision of samples with elevated $CO_2$ amount fractions, as well as extend the range of $CO_2$ amount fractions that can be shown in the results. A way to prevent that a high number of reference cylinders has to be included at all times, is to make the selection of references more dynamic. As sample measurements are always alternated with a working gas measurement, it is possible to do a 1-point calibration immediately after a sample is measured. In this way it will be possible to select the ideal set of references to calibrate the samples based on the $CO_2$ amount fractions derived from the 1-point calibration. This would save reference gas, as well as measurement time of a measurement series."*

*The combined uncertainty includes the introduced uncertainty as result of the calibration process, based on analysis of reference gases over a broad range of $CO_2$ amount fractions over a period of 2 years (Steur, 2023). Sampling biases on oxygen istotopes are studied in Steur et al. (2023), which we refer to in the manuscript. Sampling biases on the $\Delta'(^{17}O)$ specifically are discussed in the manuscript in line 227 where we state:*

*"These high differences are connected to the observations of drift in the oxygen isotopes of $CO_2$ in flask samples as a function of time (Steur et al., 2023). $\Delta'(^{17}O)$ values are not (or hardly) affected by the drifts in oxygen isotopes in the flasks. We calculated that, in the extreme case of a change of more than 3 ‰ in $\delta(^{18}O)$ of atmospheric $CO_2$ (Steur, 2023) resulting from equilibration of $CO_2$ with water inside the flask, and at the same time an initial $\Delta'(^{17}O)$ value of the $CO_2$ of -0.69 ‰, changes the $\Delta'(^{17}O)$ less than 0.06 ‰. Considering that the uncertainty of the SICAS $\Delta'(^{17}O)$ measurements is always 0.05 ‰ or higher, we can conclude that the effect of drift of the oxygen isotopes inside the flasks is negligible for the $\Delta'(^{17}O)$ values. Results and calculations that support this conclusion can be found in Appendix B1"*

2. Keeling binary-mixing analysis (and Keeling plots) is suggested to be made, to understand the endmembers, if any, controlling the variations of the isotope data. Color-coded diagram is hard to see. Scatter plots of D17O vs. d13C and D17O vs. conc(CO2) can be used to understand how much the variation in D17O is due to anthropogenic (e.g., see Liang et al., AAQR, 2017). Anthropogenic contribution (or even stratospheric influence) can also be assessed by comparing CO2 (including its isotopologues) and CO. This exercise is essential to tell whether the CO2 isotope data contain useful information, or just noise/errors from the measurements.

*Based also on a request from reviewer #1, the Lutjewad $\Delta'(^{17}O)$ summer and winter values plotted against $\delta(^{13}C)$ and $1/CO_2$ are added in Appendix D. These plots were used to show that no seasonal cycle can be detected from the Lutjewad $\Delta'(^{17}O)$ record. We would expect fossil fuel emissions to appear in the winter values as negative $\Delta'(^{17}O)$ values should correlate in this case with high $CO_2$ and low $\delta(^{13}C)$ values. However, in our dataset high $CO_2$ values are not shown due to the range in $CO_2$ amount fraction of our reference cylinders. We changed the text in line 350 accordingly:*

*"The most important difference between the Lutjewad and Mace Head $\Delta'(^{17}O)$ records is the presence of more depleted values in the Lutjewad record, with the lowest value being -0.43 ‰ in the summer of 2022. $CO_2$ equilibrated with water*
*following $\lambda_{RL}$ will have an $\Delta'(^{17}O)$ of -0.21 ‰. In summer, leaf water gets enriched in oxygen isotopes, and depleted in $\Delta'(^{17}O)$ as the result of high rates of evapotranspiration (Landais et al., 2006). Due to the active biosphere during summer, $CO_2$ and leaf water will equilibrate and the depleted $\Delta'(^{17}O)$ signal will be translated to the $CO_2$ (Adnew et al., 2023). We estimated that this could result in $\Delta'(^{17}O)$ values being up to 0.1 ‰ more depleted, when assuming the minimum $\theta$ of 0.516 for evapotranspiration (Landais et al., 2006), and considering the range of $\delta(^{18}O)$ values that were measured in our Lutjewad record. For the full estimation we refer to Appendix E. $\Delta'(^{17}O)$ values up to -0.31 ‰ can be explained by this process. $CO_2$ emitted from combustion processes has very negative $\Delta'(^{17}O)$ values (Laskar et al., 2016; Horváth et al., 2012). All points that have lower $\Delta'(^{17}O)$ than -0.3 ‰, and are sampled during winter/spring, have more depleted $\delta(^{13}C)$ values and more enriched $CO_2$ values than would be expected from the seasonal trends. This indicates that local $CO_2$ emission sources are the reason for the more depleted $\Delta'(^{17}O)$ values in winter. Samples that are very enriched in $CO_2$ amount fractions are not shown here, as that results in very high measurement uncertainties. This could be the reason that a correlation of $\Delta'(^{17}O)$ and $CO_2$ amount fractions does not appear in figure D1. A few points show depletions lower than -0.31 ‰ without $CO_2$ amount fraction enrichments, and remain for now unexplained."*

3. Need a more detail discussion on the modeling. Are the changes mainly in the D17O value in the downwelling flux or the changes are due mainly to the enhanced flux with D17O value little changed? For Eq(11), please elaborate it further. How much contribution is from the newly added 100 mbar temperature term? Is the term the anomaly from the climatology temperature? Please define "anomaly." Please compare with PV and/O3 at 100 mbar. What is the source of 0.08 per mil mentioned in Fig 7 caption? If it's from the newly added term, does it mean that the D17O from the model stratosphere is biased too high?

*In the adjusted model we add the $\Delta T$ term which will change the $\Delta'(^{17}O)$ value in the stratosphere It will, however, also have an effect on the*

*stratosphere-troposphere exchange. We argue in the text (line 416): "Note that the temperature relation represents both temperature dependence of the actual Δ'( $^{17}$O) as suggested in Wiegel et al. (2013) and the temperature dependence in stratospheric exchange, which might not be sufficiently represented with only 25 vertical layers in the current model (see e.g. Bânda et al., 2015, for the influence of vertical resolution on stratosphere-troposphere exchange)." It is easier to change the $\Delta'(^{17}O)$ production term than to adjust the air mass, as this will lead to inconsistencies in the model. The model simulation as shown in this study is therefore (still) not a correct representation of the stratospheric $\Delta^{17}O$ budget, but does show the need to include larger variability of this term in the model for a correct representation.*

*The difference between the original model simulation and the adjusted model simulation can thus be fully attributed to the added 100 hPa term.*

*The definition of "anomaly" should indeed be given and is now added to the text in line 413:*

*"The temperature anomaly $\Delta T_{100hPa}$ is determined by taking the average temperature of the months January, February and March at 100 hPa for 60-90° N per year for the period 2017-2022. Subsequently the difference between these values and the average of all 6 years is calculated."*

4. Figure 7: mid-year peak in most of the years except 2020, due to enhanced STE in spring, mentioned in the text. What is the cause of the missing peak in this particular year? Also what is source mechanism causing D17O less than 0? If I understand correctly, one has to subtract 0.08 per mil from the modified model, inconsistent with the statement -0.061-0.056 per mil variation range mentioned in Line 400. Does this mean the model was not appropriately made?

*There are two source mechanisms for an $\Delta'(_{17}O)$ of less than -0.21 ‰, as we now use a $\lambda$ of 0.528 for expression of $\Delta'(_{17}O)$, are discussed in the text. When depleted $\Delta'(_{17}O)$ values also show enriched $CO_2$ amount fractions compared to the background curve and depleted $d^{13}C$ values it is very likely to be due to fossil fuel emissions. When depleted $\Delta'(_{17}O)$ values show lower $CO_2$ amount fractions in comparison with the background curve, high rates evapotranspiration can explain the depleted $\Delta'(_{17}O)$ values. We argue that high rates of evapotranspiration can account for the values of $\Delta'(_{17}O)$ up to -0.31 ‰, when there is an active biosphere at that time. This is in the text line 350:*

*"The most important difference between the Lutjewad and Mace Head $\Delta'(_{17}O)$ records is the presence of more depleted values in the Lutjewad record, with the lowest value being -0.43 ‰ in the summer of 2022. $CO_2$*

*equilibrated with water following $\lambda_{RL}$ will have an $\Delta'(^{17}O)$ of -0.21 ‰. In summer, leaf water gets enriched in oxygen isotopes, and depleted in $\Delta'(^{17}O)$ as the result of high rates of evapotranspiration (Landais et al., 2006). Due to the active biosphere during summer, $CO_2$ and leaf water will equilibrate and the depleted $\Delta'(^{17}O)$ signal will be translated to the $CO_2$ (Adnew et al., 2023). This could result in $\Delta'(^{17}O)$ values being up to 0.1 ‰ more depleted, when assuming the minimum $\theta$ of 0.516 for evapotranspiration (Landais et al., 2006), and considering the range of $\delta(^{18}O)$ values that were measured in our Lutjewad record. $\Delta'(^{17}O)$ values up to -0.31 ‰ can therefore be explained by this process. $CO_2$ emitted from combustion processes has very negative $\Delta'(^{17}O)$ values (Laskar et al., 2016; Horváth et al., 2012). All points that have lower $\Delta'(^{17}O)$ than -0.3 ‰, and are sampled during winter/spring, have more depleted $\delta(^{13}C)$ values and more enriched $CO_2$ values than would be expected from the seasonal trends. This indicates that local $CO_2$ emission sources are the reason for the more depleted $\Delta'(^{17}O)$ values in winter. Samples that are very enriched in $CO_2$ amount fractions are not shown here, as that results in very high measurement uncertainties. This could be the reason that a correlation of $\Delta'(^{17}O)$ and $CO_2$ amount fractions does not appear in figure D1. The source mechanism for points not showing a $CO_2$ amount fraction enrichment that are more depleted than -0.31 ‰ remains unexplained."*

*The numbers that are given in the text are adjusted to the actual outcomes of the model simulation, without subtraction of 0.08 ‰, as well as these numbers are now shown in figure 7 to avoid confusion. We focus on the total variability and the timing of the peaks/throughs, as "the long-term mean values simulated by the model for Lutjewad are ultimately dependent on the integrated contribution from all processes across the globe, which are poorly constrained in the model (e.g. due to large uncertainties in soil exchange, see Wingate et al. (2009))." (line 247).*

*The adjusted text giving the total range of the measurements and the model simulation are now in line 419: "The overall variability over the full record is -0.19 to -0.07 ‰ for the model simulation and -0.27 to -0.16 ‰ for the moving average of the measurements. Although the absolute values of the measurements and the model differ by 0.08 ‰, the overall variability of the simulation with the adjusted $\Delta'(^{17}O)$ production term increased significantly and is close to the overall variability of the measurements."*

Other comments:

1. The CO2-O2 exchange method for D17O measurements was first developed by Mahata et al., not Adnew et al. Please acknowledge the previous effort.

   *The reference is added.*

2. make needed correction/clarification to small delta and big Delta in the presentation in the Introduction section.

*We revised the use of the small and big delta in the whole manuscript as, indeed, the capital delta was sometimes used incorrectly, as we should have used a small delta.*

3. Line 54: rephrase/elaborate 10 per meg for reference gas measurements. Do you mean 10 ppm is achieved for "reference" gas only?

   *The sentence is rephrased and we added another reference where the same method was used and similar precisions are reached: "The last method mentioned is at this moment acquiring a precision being better than 10 per meg for measurements of $\Delta'(^{17}O)$ (Adnew et al., 2019; Liang et al., 2023)." (line 54)*

4. Line 70: Please include Liang et al. (2023, Scientific Reports) who reported an updated data set that also include new data from Palos Verdes peninsula, CA.

   *The reference was added.*

5. Line 105-106: Rephrase/elaborate "the stability of trace gas amount fractions." It is not clear whether you referred to CRDS instrumental precision/stability or the concentrations of the gases of interest in the atmosphere/flasks.

   *Rephrased as "CRDS continuous measurements are shown as hourly means and therefore the standard deviations can vary considerably, depending on the stability of trace gas amount fractions in the atmosphere during the measurement period". (line 108)*

6. Trace gas concentration and isotope measurements: are the measurements made for the same flasks collected?

   *Rephrased the first sentence of section 2.3 (line 115) to: "Stable isotope composition measurements are conducted directly on atmospheric air samples, on the same flasks collected for the trace gas amount fraction measurements,"*

7. Line 181-186: Are the D(17O) D(17O) or d(17O)?

   *These should indeed be $\delta(^{17}O)$, this was changed in the text.*

8. Section 2.4, first paragraph. I believed you meant to compare SICAS with DI-IRMS. The first sentence seemed to say that you compared SICAS at CIO with that at IMAU. Please rephrase and make needed correction/clarification.

   *The sentence was rephrased to (line 214): " For a selection of Lutjewad samples two flasks containing identical air were sampled (from now defined as a duplo) of which one flask was measured at the CIO using laser*

*absorption spectroscopy and one at the IMAU using DI-IRMS to check the compatibility of the two methods."*

9. Section 2.4. Figure 2 caption: how is the "combined" uncertainty defined and source of errors? Is the length of the error bar 1-sigma or +/- 1-sigma? Please define it clearly. Are the errors in the difference mainly from CIO? Why the extraction at IMAU is more variable?

   *The combined uncertainty for the SICAS measurements is the same uncertainty as defined in section 2.3 (now added to the caption) and the length of the error bars is +/- 1-sigma (added to all captions). The measurement uncertainty of the IMAU measurements is considerably lower (about 0.01 ‰) and is therefore not shown. For the $\Delta^{17}O$ measurements, the majority of the points have no difference, when considering the uncertainty budget of the CIO. All points fall within the borders of ±0.05 ‰ difference, taking again into account the uncertainty budget of the CIO. For the $\delta^{13}C$ values the differences are higher and there are multiple points that fall outside the ±0.03 ‰ difference. It is hard to say what the exact cause for the differences is, as several processes can cause differences in the results from both labs: sampling procedures, storage procedures, $CO_2$ extraction procedures, as well as the different measurement methods.*

10. Figure 7: Is 0.08 per mil from the model? What's the source/cause of this?

    *We focus on the total variability and the timing of the peaks/throughs, as "the long-term mean values simulated by the model for Lutjewad are ultimately dependent on the integrated contribution from all processes across the globe, which are poorly constrained in the model (e.g. due to large uncertainties in soil exchange, see Wingate et al. (2009))." (line 252).*

11. Section 3.1 last paragraph. From Figure 3, I don't see "clearly" the drought points mentioned. Normally I'd expect drought would reduce biospheric uptake and thus cause CO2 increase. Here it said the opposite that the decrease in May-June 2018 was due to the drought. I would suggest to have a separate figure showing the deviation from the average (the background) and discuss the cause of the deviation, such as droughts, in more detail. Also there are two NOAA points next to the referred Lut(CO2) reduction, and that can be used to support the reduction.

    *We rephrased this paragraph, as the text was indeed not very clear. Also, the two NOAA points are now mentioned in the text to support the statement. Line 283: "The Europe wide drought, which was most severe in Northern Europe, during the summer of 2018 (Peters et al., 2020; Ramonet et al., 2020) is clearly visible in the continuous $CO_2$ amount fraction record of Lutjewad, where a short-term increase in $CO_2$ amount fractions interrupts the overall decrease in amount fractions that normally occurs over the growing season. In early spring of 2018, $CO_2$ amount fractions decrease*

*rapidly (when the growing conditions were more favorable, see Smith et al., (2020)), until May 2018. Subsequently a rapid increase in $CO_2$ amount fractions is observed that lasts until June, before $CO_2$ amount fractions start decreasing again. This event is only visible in one Lutjewad flask sample having a $\Delta_{bg}y(CO_2)$ of -8.6 µmol/mol and two Mace Head samples from the NOAA-GML CCGG having $\Delta_{bg}y(CO_2)$'s of -6.7 and -7.1 µmol/mol."*

12. Line 328: D17O is affected little by transpiration. It's mainly due to evaporation, or evapotranspiration.

    *Changed transpiration to evapotranspiration.*

13. Line 395: better agreement "than" the …

    *Changed to "than".*

14. Line 473: I believe here you meant d17O, not D17O.

    *We indeed meant here $\delta(^{17}O)$, which is now changed. Also we added the $\Delta'(^{17}O)$ value of the water. Line 498: "For the initial $\delta(_{18}O)$ and $\delta(_{17}O)$ of the water we use -12.91 and -6.77 ‰ VSMOW, respectively. The $\Delta'(_{17}O)$ value of the water is 0.07 ‰."*

15. Appendix B: Are the results experimental results or from model simulation? If they are experimental, please provide measurement errors? What's the D17O value of the water? With that, is the change in D17O reflected in d18O? That is, is the co-variation of d18O and D17O following water-CO2 equilibration line?

    *The results are from a model simulation, as stated in the first sentence in Appendix B (line 490): "To determine the change in $\Delta'(_{17}O)$ as the result of drift of the oxygen isotopes of atmospheric $CO_2$ inside glass sample flasks (Steur et al., 2023), a simulation of the various changes was conducted."*

    *The $\Delta'(^{17}O)$ values were added in the text in line 498 (see the answer above).*

16. Figure C1 and App C: Other than the two lowest SICAS points, there is no correlation between SICAS D17O and IRMS D17O. IRMS higher precision measurements show a factor of ~3 more variation than SICAS. IRMS as claimed has higher precision. One has to discuss whether the large variation in D17O is also seen in and supported by other data, such as CO2 (conc, d13, d18O) and CO.

    *We need to consider that we are looking at samples that were measured under different sample preparation methods. Part of the flasks that were measured by DI-IRMS at IMAU was extracted at the CIO, part of the flasks was extracted at the IMAU. Besides that, the low variance in the $\Delta'(^{17}O)$ in*

*the set of duplicate flasks, not more than 0.15 ‰, and the average combined uncertainty of the SICAS measurements of 0.07 ‰, also complicates the comparison between IMAU and CIO measurements. Figure C2 does, however, show that the general trend in $\Delta'(^{17}O)$ values measured at Lutjewad is reflected by measurements from both labs.*