Response to interactive comment of anonymous Referee #2 on "First measurements of continuous δ^{18}O-CO$_2$ with a Fourier Transform InfraRed spectrometer in Heidelberg, Germany"

We wish to thank this referee for helpful comments and suggestions for changes, which have certainly improved the revised manuscript. Our replies can be found below.

General comments:

This paper describes the extension of a previously published FTIR instrument and data analysis method (Hammer et al., AMT, 2013) to the measurement of δ^{18}O-CO$_2$. High precision measurements of stable isotopes are arguably very useful to constrain the CO$_2$ sources and sinks, and FTIR with its multicomponent capability a very attractive approach. The progress that has been achieved with respect to the precision and stability using FTIR for atmospheric measurements are impressive, and the evaluation of the method for δ^{18}O-CO2 is, therefore, highly welcome. However, to date, δ^{18}O-CO$_2$ is certainly the most challenging parameter to be measured in this FTIR setup, and it is not surprising that the original assessment by Esler et al. (2000) was very critical. While the paper by Vardag et al. is generally well written and shows promising results for δ^{18}O-CO$_2$, it lacks rigor, detail and a more critical spirit to be published in its present form. I, therefore, suggest major changes and a final decision based on the reviewed paper.

Response
We agree that the most challenging parameter to be measured with the FTIR is δ^{18}O-CO$_2$. We have expanded details of the spectral analysis as suggested by the referee in specific points below to address these concerns and provide confidence in the measurements.

1. The approach of using CO$_2$ derived from two completely different ro-vibrational bands (asym. stretching and combination mode) for isotope ratio measurement is not an established method and should thus be discussed in much more detail with respect to temperature and pressure effects as well as spectral response of the FTIR. Fig.1 is not sufficient because it doesn't reflect the real-world situation, where the FTIR is largely blind above 2320 cm-1, thus limiting the observation of δ^{18}O-CO$_2$ to half of its P-branch at best.More detailed discussion on the fitting stability, reproducibility and residual noise characterization is definitively needed. The following are some (but not exhaustive) minimal suggestions: (i) the range used for CO$_2$ (as a proxy for ^{12}CO$_2$) should be shown in Fig 1, since it is an important part of the spectroscopic evaluation, (ii) give details on all spectral regions used to quantify the different species, and (iii) show a measured spectrum in the domain where ^{12}C^{16}O^{18}O can be observed.

Response #1

We have substantially rewritten the end of section 1 and the first part of section 2 to reflect both the improvements and developments to the instrument hardware (referee point 5. below) and to provide details of the spectrum analysis. Points 1. (i), (ii), (iii), 5 and 7 have been specifically addressed and Figure 1 replaced.

2. Give more details on the cross- and interspecies correction. This is a highly critical part, well-illustrated in Fig 3 (e and f) where variations of 25 ‰ disappear yielding a surprising precision of < 0.1 ‰. For $\delta^{13}C$-CO_2, this correction is much less (< 1 ‰) and all critical parameters well discussed in the previous paper by Hammer (2013). Therefore, the authors should add a table including all sensitivities for the $\delta^{13}C$-CO_2 and $\delta^{18}O$-CO_2.

Response #2

We have added more detailed explanation (and the mathematical description) of the cross- and interspecies correction and added a table with all correction factors for both isotopologues.

3. Reproducibility is used misleadingly throughout the paper. There are several, slightly different definitions of this term (e.g. DIN ISO 21748, ASTM E- 177, GUM Annex B). However, the main (and very useful) meaning is measurements in "a set of conditions that include *different* locations, operators… on the same or similar objects" (WMO/GAW Glossary of QA/QC-Related Terminology, chapter 2.24). In this work there is no data to determine reproducibility. More specifically, the whole chapter/ title 3.2 must be adapted.

Response #3

We have checked all metrological terms in the "International vocabulary metrology-Basic and general concepts and associated terms" (JCGM, 2008). In accordance with this, we have changed the term "reproducibility" to "intermediate measurement precision" throughout the revised manuscript. Following JCGM (2008), the "intermediate measurement precision" is the closeness of agreement between measured quantity values obtained by replicate measurements on the same or similar objects under a set of conditions that includes the same measurement procedure and same location over an extended period of time, but may include other conditions involving changes.

4. The mean and SD of the difference of a number of samples (gas cylinders) is not suited to determine whether two methods are compatible (or significantly different). Use an appropriate statistical test. Change this in the abstract and the text.

5. Given the importance of this new method, the description of the instrument and the laboratory setup should be extended, i.e. referencing to previous publications is not sufficient. Also give the most important spectroscopic parameters (e.g. resolution and optical path), figures such flow, temperature and pressure stabilities, as well as the typical laboratory conditions. Explicitly state if the samples where dried (how and at which level). I assume that this instrument is a slightly modified version of a commercially available instrument (Ecotec). Acknowledge the commercial supplier and give a summary of the changes with respect to the latest version which made the $\delta^{18}O$-CO_2 measurements possible. It's important for a wide range of users to really understand why you are able to revise the original assessment by Esler et al. (2000).

6. Working gas measurements were done at a daily (24 h) basis. Since this is the time scale at which all measurements can be tight to a reference, the Allan plot (Fig. 4) should include data for 24 h.

7. In chapter 2.2.1 it is argued that the 3600 cm$_{-1}$ range has lower temperature sensitivity. However, for isotopic ratio, it is mainly the difference of this sensitivity between the isotopic species that is of importance. The authors should elaborate on this.

Response #7
This point has been addressed in the expanded section 2 describing the spectrum analysis (see also response #1).

8. Chapter 2.2.2 introduces the term "absolute calibration". Do not call this absolute because it is easily confound with spectroscopy as an absolute method. Similarly, the concept of "empirical" calibration is confusing and arbitrary. Both procedures are simple calibrations of physically sound but approximate value given by the spectrum and the fit procedure. In one case you calibrate the ratio and in the other case the individual isotopologues. Remarks: (i) I'm aware that Griffith (2012) used these terms but they are still misleading; (ii) I somewhat insist here because in the related publication by Hammer et al. (2013) implies or suggests that the FTIR values are correct (absolute), but different from the international scale (thus not absolute) "As the raw absolute mole fraction determination of the in situ FTIR analyser differs from the internationally accepted WMO scales by up to a few percent". This is not correct and not meant by Griffith et al. (2012) who correctly state that "In reality, the raw FTIR determination of trace gas concentrations is highly precise, but typically uncertain to within a few percent…"; (iii) absolute calibration is a topic in itself, e.g. attempted in detail by Griffith et al. (Analytical Chemistry, 81 (6), pp. 2227-2234, 2009).

Response #8

We agree that "absolute" and "empirical" are ambiguous in this context, and have reworded this section to make the two calibration methods clearer and have avoided the use of "absolute" and "empirical" calibration – we now use "isotopologue" and "ratio" calibration and explain the connection to Griffith et al. (2012).

9. Chapter 3.1: rewrite this paragraph using precision as a term with respect to the Allan-- deviation, and also consider less complicated phrasing to describe the corresponding results. The last sentence is misleading or at least not complete. Reaching a precision of e.g. 0.15 ‰ (30 min avg) is necessary but not sufficient to observe a diurnal cycle of 1 ‰ amplitude. What you really need is a system that is stable enough to obtain a repeatability ("…replicate measurements on the same or similar objects …", GAW glossary) over the observation period.

Response #9

Also for this term, we have checked all metrological terms in the "International vocabulary metrology- Basic and general concepts and associated terms" (JCGM, 2008). In accordance with this, we think that the Allan standard deviation describes the

measurement repeatability, which is defined as "measurement precision under a set of repeatability conditions of measurements". Therefore, we continue to use the Allan deviation as measure for the repeatability here.

We agree that the precision of 0.15‰ (30 min avg) is not sufficient to observe a diurnal cycle of 1 ‰ amplitude and that the wording might be misleading. We rephrase the sentence and state that the system is stable enough during the course of a day to significantly distinguish diurnal ambient $\delta^{18}O$ signals using half-hourly averaged measurements.

10. Chapter 3.3: These measurements are very important and as such convincing. However, it seems that the compatibility was assessed based on mean residuals of FTIR vs MS measurement results. As the target of this study is to resolve diurnal changes of delta values, the compatibility should be discussed in relation to individual flask measurements, and would be in the range of 0.5 ‰ for $\delta^{18}O$-CO_2.

Response #10

This is a critical point, which was not elaborated sufficiently in the manuscript.
If we want to observe a diurnal cycle, we need to measure accurately and compatible to the IRMS during the entire period. Every flask taken during the course of the event and measured with the IRMS must therefore compare to the simultaneous measurement with the FTIR. From the evaluation of the Allan deviation test and from the daily surveillance gas measurements, we are confident that the FTIR spectrometer is stable and precise enough to discover atmospheric signals such as the diurnal variation of $\delta^{13}C$ and $\delta^{18}O$. Also, we have shown in section 2.3 that the FTIR measures cylinders compatibly to the Heidelberg mass spectrometer. In order to show that, not only the direct cylinder measurements, but also the ambient air measurements must be compatible with the mass spectrometer. We therefore additionally compare ambient air samples, which we measure with both instruments. We have now tested the compatibility between the FTIR and the IRMS in ambient air with a two sample t-test and it was found that at the 0.01 significance level, on average the FTIR and the IRMS measurements in ambient air do not differ significantly from each other for $\delta^{13}C$ and for $\delta^{18}O$. However, we also found that the standard deviation of the difference between the FTIR and the IRMS is 0.05 ‰ for $\delta^{13}C$ and 0.42‰ for $\delta^{18}O$ and with that it is higher than expected from the Allan deviation test (Allan standard deviation of 0.03‰ for $\delta^{13}C$ and 0.25 ‰ for $\delta^{18}O$ for 9-minute averages) and the intermediate measurement precision of the IRMS (0.05-0.1‰ for $\delta^{18}O$) . The larger variability in $\delta^{18}O$ in the flask-in-situ comparison than in direct cylinder gas comparisons reflects the fact that there are more components contributing to the difference between the FTIR and the IRMS value during the ambient air comparison, i.e. the flasks itself, which could be slightly wet and then deplete the $\delta^{18}O$ value of the CO_2 in the flask, or some other possible interference of the automated flask sampler (i.e. varying integration time due to flow and pressure

variations) besides the fact that not exactly the same air is analysed by the two instruments. We therefore expect a much higher precision of individual measurements during routine ambient air measurements with the FTIR.

11. In view of the above, the discussion section should be reviewed carefully.

Response #11

We have made respective changes in the discussion.

Minor Remarks:

12. Title: reconsider the explicit statement about the place (Heidelberg, Germany) that the measurements were made. It's largely irrelevant for the content of the paper. Similarly, check whether each of the 25 entries for Heidelberg throughout the text are really necessary.

Response #12

We took the place Heidelberg out of the title and checked in the main text body for necessity of the place name. In many parts we find that mentioning the place is important, since an interpretation of the $\delta^{18}O$ signal is only feasible when a large atmospheric signal close to CO_2 sources and sinks is observed.

13. Throughout the text (including the abstract), I would suggest using the term *precision* in the context of the two-sample variance instead of repeatability. This is more common and it leaves repeatability for other circumstances, e.g. repeated measurements of a target and the derived SD.

Response #13

As described in the response # 9, we use the Allan deviation to describe the repeatability. It will not be confused with the result from repeated surveillance gas measurements anymore, since we, now, use the term "intermediate measurement precision" to describe the repeated surveillance gas measurements. This designation, as we understand, is in accordance to JCGM (2008).

14. In the abstract, the 10 min reproducibility and the 30 min Allan deviation refer to the same two-sample variance test. Therefore, use a phrase that shows this accordingly, e.g. … the precision, as derived from an Allan variance test reaches x ‰ (y min), x ‰ (y min), …

Response #14

The 10 minute reproducibility (now called intermediate measurement precision) is derived from the surveillance gas measurements, whereas the 30min Allan deviation refers to the two-sample variance test. We think that this is stated unambiguously in the abstract now.

15. Explicitly state that this is the third of a series of publications on the same instrument; Vardag et al., (2014, AMTD, doi:10.5194/acpd-14-10429-2014) is not yet cited.

Response #15

We have included the citation of Vardag et al. (2014) and set this manuscript in the context of the previous publications.

16. Step 2: it's not clear what is meant by "measurement" – at what time averaging are the corrections done?

Response #16

All corrections are done on three-minute averaged measurements. We explain this in the revised manuscript.

17. Step 2: I see no good statistical meaning of the "residual sum of squares…by the number of measurements". Even less do I see why this statistical value would be adequate to come to the conclusion "… no further concentration dependence".

Response #17

We have taken out the measure "residual sums of squares" and instead described the goodness of fit with the coefficient of determination (R^2). Further, we have stated, in the main text body as well as in the conclusion, that there is no further concentration dependence.

18. Step 4: what algorithm was used to do the smoothing?

Response #18

Daily working gas measurements are smoothed using a moving averaging filter span (points in window=10) and the residual variations are interpolated to the measurements using a cubic spline interpolation. We have added this information in the revised manuscript.

19. Chapter 3.1: cite literature for the Allan deviation; I would suggest P Werle et al., Appl. Phys. B, 57, 131–139, 1993

Response #19

We have added this citation in the revised manuscript.

20. Avoid "minutely"

Response #20

We exchange the wording "3-minutely measurement" with "3-minute averages".

21. Page 6505 Line 16: the use of δ-notation for isotopologue (and isotopomer) ratios is common for spectroscopic techniques, no need to introduce „Molec- δ^{18}O-Hitran"

Response #21

We agree that the use of the δ-notation is common for spectroscopic techniques and have changed this in the manuscript.

22. Page 6506 Line 20: Give at least some basic information on the applied mass spectrometer.

Response #22

We have added some basic information (commercial supplier, precision, scale) on the mass spectrometer, but would not like to elaborate this even more. The reader is given a citation in case more specific details on the IRMS are wanted.

23. Page 6508 Line 23: FTIR and MS measurements are not only on the same scale but also anchored via identical CO2 reference gases. Please rephrase. Is the MS an IRMS?

Response #23

We have rephrased this in the revised manuscript. The mass spectrometer used is an isotope ratio mass spectrometer (IRMS).

24. Page 6510 Line 12: Give the approximate flow applied for flushing the glass flasks.

Response #24

We have added information on the flow in the revised manuscript.

25. Page 6510 Line 7: As no results are given in Sect. 2.3 the compatibility of FTIR and MS cylinder measurements cannot be assessed … or do the authors refer to Fig. 3?

Response #25

In Section 2.3 we refer to Figure 3. This was not correctly stated in the manuscript, but in the revised manuscript we have added the reference to Figure 3 and summarized the results shown in Figure 3 in this Section.

References

Griffith, D. W. T., Deutscher, N. M., Caldow, C., Kettlewell, G.,Riggenbach, M. and Hammer, S.: A Fourier transform infrared trace gas and isotope analyser for atmospheric applications, Atmos. Meas. Tech., 5, 2481–2498,doi:10.5194/amt-5-2481-2012, 2012.

JCGM, Joint Committee for Guides in Metrology, International vocabulary of metrology - basic and general concepts and associated terms (VIM). JCGM, 2008.

Vardag, S. N., Hammer, S., O'Doherty, S., Spain, T. G., Wastine, B., Jordan, A., and Levin, I.: Comparisons of continuous atmospheric CH_4, CO_2 and N_2O measurements – results from a travelling instrument campaign at Mace Head, Atmos. Chem. Phys., 14, 8403-8418, doi:10.5194/acp-14-8403-2014, 2014.