Review of Steuer et al., Simultaneous measurement of _13C, _18O and _17O of atmospheric CO2 - Performance assessment of a dual-laser absorption spectrometer.

This paper provides a technically detailed description of a QCL laser-based analyser for the isotopic analysis of atmospheric CO_2, with a focus on calibration precision and accuracy. The increasing availability and usage of optical methods (as distinct from isotope-ratio mass-spectrometric methods, IRMS) for isotopic analysis of trace gases in air makes this study both timely and valuable. Optical methods provide many advantages for isotopic analysis not easily accessible by IRMS, such as minimal sample-preparation and continuous analysis in whole air at sub-minute timescales. However optical methods currently have lower precision and accuracy than IRMS, approaching but not meeting the compatibility requirements set out by WMO's Global Atmosphere Watch (GAW). This paper provides a detailed assessment of a state of the art laser spectrometer for precision and accuracy using two different approaches to calibration. It is clearly suitable, valuable and desirable for publication in AMT, but I do have some reservations about the detailed treatment of the calibration schemes as described in the submitted paper and suggest major revisions before publication. I also provide technical comments. When addressed I believe both will improve the paper substantially.

General comments

General comments and suggestions for revision are centred around the comparison between the ratio method (RM) and isotopologue method (IM) for the calibration strategy. These methods have been discussed and dissected in detail in the past few years, culminating at the 2017 GAW GGMT meeting in Dubendorf. Several papers from that time and earlier have now been published but, given the central focus of this paper on the calibration methods, they are not well presented or critically reviewed here. In particular for the more novel IM, papers by Wehr et al (Ag. For Met 2013, including authors common to this paper) Tans et al (AMT 2017) and Griffith, (AMT 2018) are not cited or discussed. (Disclosure: The reviewer is the author of Griffith, AMT 2018. See that paper for references to these and other relevant papers.) The analysis schemes used in those studies are equivalent to that of Flores et al (Anal Chem 2017 and Griffith 2012) which are referenced here. Griffith 2018 supplemented and extended Flores et al and included a review and analysis of the two calibration strategies. Given the central focus of ratio vs isotopologue calibration in this paper, I recommend adding a short but critical review (or at least a comprehensive literature survey) of existing work on calibration methods as they relate to the present work.

The paper is generally technically well-detailed and comprehensive in describing the analysis, after addressing technical comments and suggestions listed separately below. The major exception, based on the presented material, is the treatment of the isotopologue method (IM) for calibration. Since both RM and IM use the same raw measured data with the same uncertainties and noise characteristics, I would intuitively expect both calibration methods in principle to deliver the same accuracy and precision if applied with equal rigour. So I ask why the accuracy and precision as presented is significantly lower for IM than RM? I think the answer lies in the detailed application of the two treatments, with and without consideration of drift, non-linearity and concentration dependence in the calibrations, as described by Wen et al (2013, referenced):

- Non-linearity, due to imperfect spectroscopy and spectral fitting (eg imperfect knowledge of line shapes, pressure and temperature dependence, approximations in the spectrum model used for the spectra). Non- linearity is small, it applies to both RM and IM, but is only dealt with for the RM, not IM. The section on concentration dependence (3.1) quantifies the non-linearity, but in the calibration section 3.3 non linearity is only considered and dealt with for RM, while IM is based on simple 2-point linear calibration for each isotopologue.

- Non-zero offset in the individual instrument response to each isotopologue. This is included explicitly in the linear calibrations for IM, and implicitly through the empirical non-linearity / concentration dependence correction in RM.
- ~~Following Griffith 2018, non-linearity (expressed as a quadratic term in the relation between instrument response and mole fraction) leads to a pseudo linear concentration dependence on raw isotope ratios, and a non zero intercept leads to a pseudo-inverse concentration dependence.~~

In RM, the sample is measured alternately with WG and QC gases, emulating practice in IRMS. The WG and QC gases provide both an interleaved constant reference R, eq. 5 and 6, as well as to derive and apply a real-time empirical concentration dependence, eq. 9, for each measurement sequence. This scheme appears to work well, taking account of both instrument drift and changing characteristics, as well as CO2 dependence and non-linearity, continuously.

In IM, however, the description provided is much less detailed but from what is presented it appears that only a single linear 2-point calibration, determined once for each isotopologue, is applied to relate the instrument response (Xm in eq. 10) to the actual isotopologue amount (Xa). This ignores both non linearity and any change or drift in the calibration coefficients c and d with time, both of which are implicitly corrected in the RM approach. The root cause of any change in calibration with time is the same in both methods – it is embodied in the non zero intercept and quadratic terms in the effective calibration relationship between instrument response and actual isotopologue amount. As this section 3.3.2 is far less detailed than 3.3.1 and I have less information to work with, I invite the authors to correct me if I am wrong, but in any case extend the description of the method for IM in 3.3.2 to address this lack of detail.

I suspect this unequal treatment of calibration changes with time may be the root cause of the different precisions and accuracies found for the two methods in measurements of flasks and other tested samples. The measurements are of good quality, extensive and time consuming and presumably cannot be repeated easily. But I ask if the existing co-incident WG and QC measurements collected in each measurement run could be used to effectively re-derive or correct the coefficients c and d in eq 10 in real time to address stability and drift, as they are for RM through the R gas and CO_2 dependence corrections.

Until this is resolved, in the paper as written the comparison between RM and IM is not an equal one. When addressed, this may (or may not) change the final conclusions from the paper. I suspect that the precision and accuracy for each method applied with equal rigour may be quite similar. The discussion would then centre around the relative measurement effort required by each method to achieve the required precision and accuracy rather than a simple one-is-better-than-the-other conclusion.

In the interest in providing an on-time review, I do not provide comments on the results, discussion and conclusions sections since they may change substantially.

Technical comments

Abstract: In the interest of readability, I think best editorial practice is to avoid abbreviations in the abstract, and to introduce them at first use in the main text. I agree with reviewer 1 that there is some over-use of abbreviations and acronyms. Depending on the response to the general comments above, the key conclusions from the study may change.

L27: Low variability is due to the large size of the carbon reservoir and the long lifetime of CO2 in the atmosphere, not the high mole fraction. (Note: mole fraction is strictly the correct term, not mixing ratio)

L31: Can this reference to WMO 2016 be updated to the latest (from 2019 meeting).

L49: Optical methods to which this paper applies include non-laser methods such as FTIR and even NDIR. Remove (laser), it is in any case appropriately mentioned later in the sentence.

Section 1 General – see general comments, please include a literature survey and preferably some critical analysis of existing work in the calibration topic.

L65: replace “consists (among others of)” with “includes”

L68: In contrast to Reviewer #1 I am comfortable with the Hitran notation – it is well established (in optical spectroscopy at least) and quicker to type and read than full molecular labels, especially in equations.

L69: (with a sweep frequency of…) Add sweep to avoid potential confusion with optical frequency.

L74: Please specify that this is the path inside the cell (the outside path is dealt with in the next paragraph, but it is ambiguous here)

L79: How accurate is the temperature control? This is relevant in assessing the causes of drift which are important to the whole calibration strategies. Likewise pressure control (see line 201).

Figure 3: requires a full caption with all labels and abbreviations spelt out. It is quite difficult to follow the text on p5 with the current level of detail provided. L91 – where is the inlet valve, for example?

L98: mbar not mBar, and please specify this is gauge pressure, not absolute.

L109 and Fig 4. What are the units of the Y axes of plots? For the upper plots they should be ppm^2 but I cannot relate this to the mimimum figures given in the in-plot tables. The lower plots are presumably ‰^{2} . Also the abbreviation “St. dev”; these are not standard deviations, they are the minimum Allan deviations (ie sqrt(AllanVariance)), which is strictly not quite the same thing. Finally, unless the amounts for the isotopologues are scaled by their natural abundances, the Allan variances in ppm^2 will have quite different absolute values.

L143: please replace “n” with “number of measurements”

L143: should “lower” read “higher”?

L129: what is “demi-water”

L152: This is not a fair comparison, or is a misleading statement. It is only to be expected that ratios will show lower Allan Variance than individual isotopologues because some sources of noise and drift in the ratioed quantities are correlated and cancel. When isotopologue amounts are ratioed later in a calibration calculation, as in the IM, these correlated noise sources will also cancel. This should be made clear.

Table 1: These are relative standard deviations in ‰, please add “relative” to the caption and make clear in the case of the ratios that it is not the delta value itself but the relative std dev of it.

L180, eq. 3. Although this equation is only used for illustrative purposes, it is confused by the true calibration equation (10). The difference is between X_{CO2} and X_a. X_{CO2} is itself a function of X_* (or X_m) through the relevant isotope ratios so the coefficients *a* here and *c* in Eq 10 are different. It would be clearer to use the same equation as (10) here.

L214: Griffith2018 showed that a pseudo linear + inverse relationship to CO_2 amount is both theoretically expected and fits data in that paper in practice. The quadratic fit is only an approximation to this.

Section 3.2.1 and Fig 5. I have difficulty following this section and figure. It may be better titled “Spectroscopic non-linearities of isotope measurements”. If the “residuals” are the differences between a linear regression of X_* against X626, they should be in units of ppm, not ‰ in Fig 5. I do not understand the sentence on L 223 “The CO2 mole fraction…”

L249: There should be no apostrophe in “deltas”. An apostrophe is used to indicate possessive, not plural. There are many examples through the MS.

Fig 6. See comment for L214, it would be interesting to see these δ data plotted against X_{CO2} and $1/X_{CO2}$. Does this improve the comparisons in Table 3? The description of this table is also difficult to follow in the text.

L245: Please clarify the sentence “From now on we use the δ with an isotopologue superscript….” and/or give an example. It is quite unclear to me.

L256-283: This section is important in the context of general comments above on correcting for stability, non linearity and drift. It points out that the calibration changes with time, but my understanding of the analysis is that this is compensated in RM but not IM. This might be clarified in dealing with earlier comments on the IM in 3.3.2.

3.3.2 Isotopologue method. Please see earlier general comments on this section. This does not have the same level of detail as 3.3.1 and if I understand it correctly the IM does not include the equivalent of the interleaved comparison with a reference gas R, or any correction for changes in the coefficients c and d in Eq 10. These changes are the root of any calibration drift in both RM and IM . As this section is written they appear to be corrected in RM but not IM.

4, 5, Results, discussion and conclusions. I do not comment on these sections because they may change substantially if the comments around the treatment of IM can be addressed.