

Reviewer comments

Malina et al., On the consistency of methane isotopologue retrievals using TCCON and multiple spectroscopic databases

General comments

This paper assesses comparative retrievals of both $^{12}\text{CH}_4$ and $^{13}\text{CH}_4$ total column amounts from TCCON spectra, with an emphasis on a new spectral window and line parameter data available for the TROPOMI instrument on ESA satellites (S5P, S5), but not currently used by TCCON. The study covers 4 separate line parameter lists and 4 spectral windows, with spectra of differing quality from 2 TCCON sites, Tsukuba and Ascension Island.

The study is thorough and detailed, covering the sensitivity of $4 \times 4 \times 2 = 32$ different combinations of analysis to a number of potential sources of error. In general I have no argument with the thrust and rigour of the work, apart from specific comments below. However I find the level of detail difficult to assimilate, so that the key conclusions get lost in the (repetitive) detail and length of the paper. I would make the general recommendation to move much of the detail expressed in similar arrays of plots and figures to supplementary material or an appendix, while providing an essential summary of each comparison and sensitivity test in the main paper. I do not provide a detailed roadmap for this, but I have made some suggestions in the specific comments. This would shorten the main paper and improve its readability for a wide audience, while not removing the detail for those readers who need it. The conclusions do provide a reasonable summary of the main points from the study, but the road is long to get there.

I have made comments below about the feasibility of making *useful* remote sensing measurements of $\delta^{13}\text{C}$ in CH_4 . I agree with the authors' final conclusion that this is not currently feasible with sufficient accuracy to be useful, but this part of the analysis is over emphasised through the paper because it is clear from the start that this is the case. The sections on $^{13}\text{CH}_4$ could be much reduced while still coming to the same valid conclusion.

The paper is suitable for publication in AMT, but I recommend a major revision to restructure it to bring out the key messages from the large amount of detail.

Specific comments

P1 L15: Results obtained with the SEOM-IAS database *in the 4190-4349 cm^{-1} window* show the lowest fit residuals

P2L40: Can you provide a reason why TCCON uses 6000 cm^{-1} and does not use this window? Is it because of the greater interference from water vapour? See also P3 L60.

P2L52: My understanding is that solar occultation refers to measurements from space (or at least above most of the atmosphere) where the ray path follows the sun as it sets or rises through the earth's limb between high elevation angle (exo-atmosphere) to low. TCCON would be better described as taking direct-beam solar absorption measurements from the ground.

P3L82: $^{12}\text{CH}_4$ is closer to 99% than 98%. This paragraph overstates the possibility of retrieving useful information on $\delta^{13}\text{C}$. As pointed out later (p8, discussion and conclusion), $\delta^{13}\text{C}$ in CH_4 varies only by around 1‰ from -47‰ *in situ* near the ground except close to strong sources, and we would expect total column amounts to vary much less than that. The possibility of measuring total column $\delta^{13}\text{C}$ in CH_4 with sufficient accuracy and precision to be useful for source attribution is not just challenging, it

is very remote indeed. Although the authors do mention the challenge, this paragraph holds out a false hope that could have been recognised from the beginning.

P4L99 Remove “While” – not a sentence. (There are other instances...)

P4L123: This is ambiguously written – GFIT assumes a fixed profile shape for EACH trace gas – but they are different for each trace gas

P4L130: most spectral **features** are resolved, but many spectral **lines** will overlap at the pressures and temperatures in the atmosphere.

P7 L188: Should σ (intra-window) rather read σ (inter-window)?

P7L192: A posteriori error appears twice.

P7L196: the denominator should read $(^{13}\text{CH}_4/^{12}\text{CH}_4)_{\text{VPDB}}$ not just VPDB

P8L210: Using CO accuracy and precision requirements as a proxy for $^{13}\text{CH}_4$ because it has a similar total column amount is misleading and irrelevant. $^{13}\text{CH}_4$ measured with the target precision and accuracy for CO is geophysically/geochemically useless. We already know the value better than that -it is very closely defined by the $^{12}\text{CH}_4$ column because $\delta^{13}\text{C}$ varies by less than 1‰ and $^{13}\text{CH}_4$ can be calculated from Eq. 2. What really matters is how much $^{13}\text{CH}_4$ varies from this value. That requires far higher precision and accuracy. The text and objectives for $^{13}\text{CH}_4$ measurement should reflect this reality in this and the paragraph beginning on L219.

P8L218: See comments above. this quoted uncertainty of <1‰ is required in in situ measurements near the surface – once the small variations near sources of varying $\delta^{13}\text{C}$ are mixed into the (almost unvarying) total column by the retrieval, the requirement is <<<1‰.

P9L232: There is ambiguity between the line parameters, which are intrinsic properties of the molecule, and the spectral lines as observed and analysed, which depend on both the intrinsic line parameters and the external conditions – concentrations, pressure, temperature. The databases differ most likely in line positions, strengths, lower state energies and widths, as well as completeness – some lines may be missing in one list compared to another. These parameters are not, or are only weakly, pressure and temperature dependent themselves – they determine the positions, strengths and widths of absorption lines and their temperature- and pressure-dependences. – eg. the lower state energy is the main determinant of temperature dependence of a transition and hence absorption strength – it is not temperature dependent itself.

The max/min/mean of line intensities and lower state energies seems like a very crude measure for meaningful comparison, and I do not find 3.5.1 on P23 very helpful. This section (3.5.1) could be put in the appendix and the main conclusion summarised in a much briefer section.

P10L270 section 2.4.3. I suggest deleting this section. The case that a different apodising function is used for measured and modelled spectra would be a gross error in the analysis and should never occur. If apodisation is applied in GFIT, it either matches that applied to the measured spectrum, or is applied internally to both and there is no mechanism by which it can be applied to one and not the other. There is little point in evaluating it.

On the other hand, there may be (and are) small instrument effects which make the ILS imperfect, and which could be viewed in a similar way to apodisation. Such effects should be much smaller than the difference between an apodised and unapodised spectrum, even for a weak apodisation function. A better ballpark measure of the effect of imperfect ILS would be to slightly adjust the instrument Field of View or phase error, which affect the width and symmetry of lines.

P12 Fig 2: This figure has limited value. It is impossible to distinguish the differences between the 4 databases through the four coloured traces, either through congestion, or poor choice of Y-axis scale. The most important information is in the different residuals (row 3). I suggest that the 5 single component spectra in each region, appropriately and comparatively Y-scaled to be easy to view, be moved to supplementary material/appendix or dropped, and Fig 2 replaced with only the CH₄ spectra (4 columns), perhaps H₂O as the major interferent, and one row each for the 4 residuals, having the same Y-scale in each window for comparability. The differences between databases would be much clearer in this view, supported by the statistics of Table 3. Ascension Is spectra do not need a figure in the appendix, comparable statistics to Table 3 would suffice (with a reason for the higher noise/poorer fits).

P16 Table 4. Rows 4 and 5 should read inter-window – see comment P7L188. If I understand correctly, this is the deviation across all windows, (intra-window would mean variation within one window, which seems meaningless).
From line 7, replace “b” with “bias”.

Figure 5 could be removed to the appendix and just the essential differences to Tsukuba outlined in the text here. What is the reason for the higher uncertainties at Ascension?

Table 6 nicely illustrates the difficulty of retrieving useful information on $\delta^{13}\text{C}$ in CH₄, given that we know a priori that the correct answer should be within 1‰ of -47‰. See earlier comments.

P24 3.5.2 I found this section long and difficult to assimilate. In Fig 9, what are the dotted lines? And in Fig 10, it is very hard to visualise or interpret the differences when both slope and intercept are varying and are co-dependent. It is much easier to interpret comparisons of a single metric, even if less rigorous. For example if the data of Figure 9 included the 1:1 line it would be much clearer to see any offset and slope differences and make an immediate comparative visual assessment. Fig 10 would not be necessary, but could be replaced by a table of comparative single statistics. Likewise Figs 11/12, 13/14 and 15/16.

In the case of the 2% CH₄ bias, is this just a scaling of 2% to the a priori profile? I would not expect a large sensitivity in this case, because the retrieval also simply scales the a priori total profile without changing shape – only the initial starting point changes. Perhaps a better test would be to perturb the shape of the profile to assess the sensitivity, by changing the tropopause height or the tropospheric/stratospheric balance.

Sensitivity to water vapour is perhaps the most important here because water vapour is so variable.

P31 3.5.3. See earlier comment, I do not think this ILS sensitivity assessment is meaningful in practice in this form.

P35 L521: Long term averaging of ¹³CH₄ measurements at a single TCCON site may well increase precision, but accuracy is more important and is not improved by averaging. Table 6 illustrates that the accuracy needs to improve by 1-2 orders of magnitude to be useful.