

Reply to Referees #1, #2 and #3

We would like to thank all three referees for their very helpful comments and suggestions which will improve our manuscript. We provide here explicit responses to all reviewer comments, and attach a draft revised manuscript incorporating these responses. In the following, the referee comments are italicized in black, followed by our replies in red.

Anonymous Referee #1

Received and published: 10 July 2012

General Comments

• *I find the Conclusions and Outlook section addressing the calibration gas consumption and recommendations for number of frequency of calibrations to be particularly useful for users of the instrument. It would be interesting to contrast this to the needs for other instruments that adhere to the WMO-GAW criteria. It would also be nice to recap on the perceived frequency necessary for performing the cross-sensitivity quantification experiments, which is mentioned elsewhere.*

We included a discussion on the other optical techniques in the conclusion and outline section. In particular we added a sentence on the lifetime of the calibration gases:

“Although the lifetime of a FTIR calibration gas cylinder is thus longer than for classical GC systems (here the typical lifetime is one year), it is shorter than the expected calibration gas cylinder lifetime for other optical techniques like Quantum Cascade Lasers (QCL) or CRDS. For the latter technique the calibration cylinder lifetime is expected to be on the order of decades, depending on the calibration scheme (e.g. Winderlich et al., 2010).”

• *At the end of Sect. 3.3, for example, it is stated that the ILC targets are considered to determine whether the sensitivity is significant. While that may be true over the typical range of variation quoted, how do these effects cumulate? Does that mean that the effects for these gases are ignored, or are the corrections still applied?*

We apply all residual- and cross sensitivity corrections even after we have changed the hardware to reduce the cause of the variation, making some of the residual corrections marginal.

• *It would be nice to include a table listing the typical variability of sample conditions (temperature, pressure, flow...) for easy reference. This could also list the standard operating conditions (c.f. Sect. 2.3).*

The first column of Table 1 lists typical variability of the sample conditions..Thus we refrain from adding another Table.

• *p3660, L25-27: why would the sample pre-treatment compensate for the temperature difference? Is it the treatment itself, or the residence time in the dryers? If the Nafion is housed outside the thermo-controlled enclosure can this be relied upon to compensate for temperature differences in reality? How important is the role of the laboratory conditions on the temperature stability introduced by the Nafion?*

The temperature compensation is most likely accomplished by having sufficient residence time in the Nafion. The Nafion itself is not located in the temperature controlled compartment; it is located in the lowest compartment housing with the tubing, the electronics and the power supply. Temperature in the lowest compartment is around 60°C. The laboratory temperature has only very little influence on the temperature in the lowest compartment. Furthermore, since the Nafion is operated in counter-flow mode we believe that the heat transfer takes place between the outflow and the inflow gas sample. The sample inlet temperature compensation effect of the Nafion can thus be assumed to be relied on.

For a better understanding we changed the text to:

“This implies that the residence time of the sample in the drying system, i.e. the Nafion dryer, is sufficiently long to compensate for at least a 30°C temperature difference of the incoming sample.

• p3662, L18: What is the lifetime of one magnesium perchlorate cartridge? Is it a recommendation of this paper to maintain the H₂O between 2 and 10 $\mu\text{mol mol}^{-1}$? Otherwise what governs the operational range, and therefore the range that is examined here? Do the authors have any idea how the cross-sensitivity with respect to H₂O behaves outside this range that might feed through to making a recommendation for the operational H₂O limits? At what point (i.e. H₂O mole fraction) does the cross-sensitivity to H₂O start to limit the ability of the instrument to reach the WMO-GAW standards?

The lifetime of one magnesium perchlorate cartridge is typically 2 months. We have included this information in the text. Yes, we recommend keeping the moisture level below 10 ppm. The wider H₂O range is explicitly discussed and shown in Figure 4, and the H₂O cross sensitivity cannot be easily corrected in the wider range since the hysteresis effects come into play. At the end of the lifetime of the drying cartridge the moisture level increases typically from 2 to 3 ppm within four days. The moisture increase is exponential. Still this slow start in the increase allows for enough reaction time to change the cartridge.

• While I appreciate that the linearity of the instrument has, in general, been established, I would expect that three calibration standards are necessary to adequately define the line and/or assess the linearity. After all, one can define a line with only two points - if one of those happens to be incorrect then your line will be misdefined. I would think that some redundancy is necessary. The conclusions mention this, but it could also be addressed earlier (end of Sect. 4).

We agree that for high accuracy demands the usage of more than two calibration standards is necessary. Thus, we have included a respective sentence at the end of section 4:

“For high accuracy demands, as in atmospheric background monitoring programs, the usage of three calibration standards to define the IRF is advised. A three point calibration reduces the sensitivity to individual outliers and instrument noise in the calibration measurements.”

• p3667, L25-28: Do the authors know why there is disagreement between the Griffith Allan variance results and those presented here? Is it instrument specific? Is this important?

The results of the Allan variance are instrument specific. Our experience is that the time until the Allan variance improves depends upon the stability of the sample parameters like temperature, pressure and flow. The Allan variance is thus not primarily limited by the spectrometer itself; rather it is caused by the gas handling. Since the IUP configuration of the FTIR improved these sample parameters, the results of the Allan variance experiment could be improved as well.

Technical Comments

• Abstract: There is an inconsistency about whether or not acronyms are defined - e.g. ILC is, but WMO-GAW, FTIR are not.
All acronyms are now defined

• Introduction: p3646, L21 (and numerous other places) - the use of which is incorrect. Please check this throughout the manuscript. As a rule, ‘which’ should either be preceded by a comma or a preposition. In many cases it could be replaced by ‘that’, requiring no preceding comma. The following is an explanation of the difference between using ‘which’ and ‘that’.

– “The painting, which was hanging in the hall, was stolen”

– “The painting that was hanging in the hall was stolen”

Both mean that a painting hanging in the hall was stolen. The former simply offers a property of the painting - this does not mean that it was the only painting in the hall. The latter, on the other hand offers a unique description identifying the painting - it was the one hanging in the hall. Another exception occurs when addressing a question, e.g. when answering the question: “Which calibration strategy is better?” one could reply “We could not determine

which calibration strategy was better' There are numerous instances of the misuse of the word 'which' that I will not specifically mention.

Thank you very much for your patience in explaining English grammar. We tried our best to prevent the misuse of the word 'which'.

- p3648, L23: Further on ! Hereafter?

Done

- p3648, L27: The wording at the end of this sentence sounds a bit awkward. I suggest rephrasing to "the hardware modifications resulting from the first year's findings are also introduced" (note the plural with modifications).

Done

- p3649, L7: what does the (5.0) mean when referring to the high purity nitrogen? For those not familiar with the grades, it would be nice to list the purity (99.999% ?) or at least refer to it as 'grade 5.0'.

Included: purity: 99.999

- p3649, L15 c.f. p3650, L11: Please be consistent with the spelling of analyser/ analyzer.

Done

- p3649, L17: Would be nice to have the name (magnesium perchlorate) as well as the chemical formula.

Included

- p3649, L18: either by-passed or bypassed, but not by passed.

Changed to: by-passed

- p3651, L3 and L23: I think the 'the' before SOC is unnecessary.

Removed

- p3651, L3: suggest replacing 'compartment' with 'enclosure' for consistency

Accepted

- p3651, L22: Delete 'on' at the end of the sentence.

Done

- p3652, L3: ability of completely exchanging the sample ! ability to completely exchange the sample OR ability for the sample to be completely exchanged

Changed to: completely exchange the sample

- p3652, L10: results in 0.03 ! results in a 0.03

Done

- p3653, L3: minute ! minutes

Done

- p3653, L19: which ! that

Done

- p3654, L1-4: How valid is this assumption? It would be nice to at least refer to the fact that this assumption will be addressed in later sections.

As long as it is a systematic, concentration- and sample type -independent error, it will be corrected for by the calibration. We could show that for the UoW and the ICOS Demonstration Experiment setup the sample types have not been comparable which lead to the TDS. However, in the improved final IUP-setup we could overcome this problem.

- p3655, L15: delete 'do'

Done

- p3656, L2: 'cylinder filling' - while I can appreciate what this means, it could be better expressed. Perhaps 'Each time a cylinder was filled, it was checked for drifts by GC analysis before and after use', or something similar.

The GHG concentrations in each test cylinder were checked for drifts by GC analysis before and after use.

- p3656, L10-11: ?? This sentence requires rephrasing for clarity.

A residual or cross sensitivity experiment determines the concentrations of a fixed sample with respect to at least four different settings of the investigated sensitivity parameter.

- p3656, L17: As reference ! As a reference

Done

- p3656, L23: Incorrect English. Suggest rephrasing to '0.1 or 0.05 $\mu\text{mol mol}^{-1}$, respectively, for CO₂ in the Northern and Southern Hemispheres.'

Done

- p3656, L25: repetition of 'determined' - delete one instance.

Done

- p3657, L3-4: The way this is phrased is confusing, given that the second range is within the first. I suggest rephrasing to state that the second range, or a subrange, (around operating pressure) is tested in finer pressure steps.

Two different pressure ranges, one from 800 to 1200 hPa and a sub-range around the operating pressure from 1085 to 1115 hPa, were tested.

- p3657, L7: 'residual pressure sensitivity' - you've just gone to the trouble of defining this acronym, why not use it?!

Changed

- p3658, L14: later used thermocouple ! thermocouple used later ?

Section changed

- p3658, L15: at least on ! on at least

Changed

- p3658, L21: what is 'the whole system'?

Changed to: spectrometer and the sample cell

- p3658, L27: move 'was' to after 'experiments'

Done

- p3659, L24-26: You choose a linear fit over a step change. Is there any physical reason to prefer this?

With the newly included discussion on the origin of the TDS the assumption of a linear fit can directly be deduced from equation 1.

- p3660, L9-10: was the sample flowing or static after the evacuation?

Included "in flow mode" for clarification:

"While in the dedicated RTS experiments sample air was continuously flushed through the cell, each cylinder measurement, performed under standard operating conditions, comprises evacuation of the cell prior to the measurement in flow mode (see section 2.3)."

- p3660, L27-29: investigations ... is ! investigation ... is OR investigations ... are

Corrected

- p3662, L7-8: differently pronounced ! shows different behaviour?

A further challenge is a hysteresis effect in the H₂O cross-sensitivity which is differently pronounced for each individual species

Changed to:

A further challenge is a species-dependent hysteresis effect in the H₂O cross-sensitivity.

• p3662, L15-16: suggest rephrasing this sentence to: 'They may, however, become relevant if water and water isotopologues are to also be measured with the instrument.' It might be nice to have a reference here too, because the fact that the instrument can potentially perform H₂O and H₂O-isotopologue measurements has not been introduced in this paper, unless I have missed it. I notice that it is mentioned in passing earlier (p3649, L19) - it could also be expanded there, maybe just a part sentence saying that this is possible with the instrument, and referring to a publication with details.

Changed the sentence and added a reference to the earlier passage.

• p3663, L3: means ! methods

Changed

• p3663, L24: what about H₂O production from the Ascarite, via the 2NaOH + CO₂ ! H₂O + Na₂CO₃ reaction? Do you see any effect from this, or counteract this is any way?

Added this sentence for clarification:

The H₂O that is produced by Ascarite during the CO₂ uptake is removed by the drying system.

• p3665, L11: This sentence needs rephrasing to avoid the misuse of respectively.

Seven or five of these cylinders have also been calibrated for N₂O and CH₄ by NOAA/ESRL, respectively.

• p3667, L1: the (laboratory?) temperature oscillated...

Although our laboratory is air-conditioned, the temperature oscillated between 23.2 °C and 25.1 °C.

Changed to: The Heidelberg laboratory temperature oscillated between 23.2°C and 25.1°C.

• p3667, L20-22: You previously discussed the choice of timestep (p3651, L15), so this is both redundant, but also raises other reasons from what was mentioned earlier. I suggest mentioning all reasons (or removing those not applicable) in one place, and one place only.

Remove this part and extended the discussion in section2.

• p3668, L16-17: What question about calibration frequency?

In the latter case the question about the frequency at which the calibration need to be performed arises

• p3668, L21 (and repeated afterwards): suggest calling these 24- and 48-hourly. To me two-daily sounds strange.

Changed

• p3669, L22: regular, e.g. daily calibration might help reducing... ! regular, e.g. daily, calibration might help to reduce..

Changed

• p3670, L1: What does 'One outlier appeared due to a wrong calibration measurement' mean? Was the wrong tank measured? Was the value simply an outlier? Please explain.

We changed this sentence to be more precise:

Outliers in the target gas record (July 2011) are caused by bad calibration measurements.

• p3670, L23: define MPI-BGC

Done

• p3670, L26: replace mid with middle (both instances)

Done

• p3671, L5-6: Which step change? Suggest that you explain with reference to the fig when this occurred.

Included the date and a reference to the Fig.

The decrease in reproducibility for CO is caused by the step change on July 15th (compare Figure 9) and for N₂O by several outliers.

- p3671, L17: Suggest that you define the periods that are considered 'stable' and 'less stable'. Perhaps with a horizontal bar in Fig 10 corresponding to one of these.

Done

- p3672, L10: by a factor 2 ! by a factor of 2. Should this actually be $\sqrt{5}$? In the following line, should this be 100% larger, as you are calculating relative to the target measurements?

Yes, you are right! Thanks. Changed the text accordingly

- p3672, L12-13: Different from other IRFs ! Different from the IRF for other gases??? (just for clarification, as opposed to the possibility that this might refer to other instruments)

Changed to: Different to the IRF of other gases, the IRF for N₂O has two large step changes during the stable instrument conditions (not shown)

- p3673, L16: not observed anymore ! no longer observed.

Done

- p3673, L17: 0.1_ only ! only 0.1_

Done

- p3674, L20: not sure about the wording in brackets

Changed to: (refer to section 2.4)

- p3675, L19: formally not ! not formally

Done

- p3675, L25: a few $\mu\text{mol mol}^{-1}$ of H₂O, only ! only a few ...

Done

- p3676, L18: left over ! remaining

Done

- p3676, L22: to use ! using (or 'the use of')

Done

Other: it would be nice to acronyms defined in the figures when they are used.

Done

Fig. 1: is it possible to have arrows defining the direction of gas flow?

Done

Fig. 7: our? primary laboratory cylinders.

Changed to: of the secondary laboratory cylinders used for the Heidelberg GC system

Fig. 8: it is really hard to see the grey points behind the coloured symbols. It would be nice if something that stands out more could be used.

The grey points can only be seen for CO₂. For all other species the differences between the residual- and cross-sensitivity corrected results are marginal and in this plot invisible. We added a respective statement in the figure legend.

Fig. 9: some residual German (Mai) in the x-axis label!

Removed

Anonymous Referee #2

General Comments:

I find the abstract to be extremely thin and lacking in detail. The paper is a laborious read and important details will be difficult to extract for a reader glancing through its content. I can appreciate that these findings and the associated specific details are important for inclusion in an “instrumental type” manuscript but it would be handy for an interested reader to find a summary of important findings to spark his/her interest in obtaining such an instrument. These results will be of great interest to parties linked with surface GHG observational programs. The abstract should include short statements on specific findings of this study. Statements such as “moderate” stable laboratory conditions (L11) are too general.

We have revisited the abstract and included the (quantitative) key findings of the paper.

Although I noted that the overall lay out of the presentation is fine, it’s not ideal. The structure is not framed to get the important messages out. Many of the interested facts are masked by descriptive sentences and background information. The paper should be framed for the audience that it is intended for. A more appropriate layout would have included many smaller sections highlighting results with further reference to additional information in auxiliary sections. In reading the paper, I found the figures and tables to be most useful in providing details, not the text.

We have reorganized the paper and introduced more sections and moved less important information to the Appendix.

Lastly, as an interested reader, I was left wondering how this instrument compares to other commercially available instruments such as Cavity Ring Down Analyzers. There was mention of benefits of the FTIR over the use of gas chromatographs. For example, the possibility of measuring 5 species with one analyzer is of great interest and the precision of the ^{13}C is most impressive. Although the ^{13}C may not meet WMO comparability requirements, the continuous mode of sampling and precision is attractive for use in applications where strong variations of the ^{13}C , such as in forested sites and urban centres, are high. I can appreciate that comparisons to other instruments may be out of the scope of this paper, but none the less, even statements that the precision levels are in line with other current insitu commercially available instruments might be important to include.

We included a reference to other optical state of the art GHG analyzers in the outline of the paper. A comprehensive comparison between the different analyzers is however, as you mentioned, out of the scope of this manuscript.

V. Sherlock (Referee #3)

General comments

It was a pleasure to review this paper. While entirely technical in scope, it represents an exhaustive study of the performance of the FTIR trace gas analyser developed by the University of Wollongong and now commercialised by Ecotech. The paper is dense in detail, but I am sure it will serve and be widely cited as a leading reference for FTIR analyser measurement and calibration procedures for some time to come. I do have some suggestions which I would encourage the authors to consider/implement to assist a general reader to access the information in the paper more efficiently:

- *A short section providing a brief overview of the sample data acquisition, analysis (retrieval through cross-sensitivity correction) and the calibration procedure should be added immediately after the introduction. This should give the reader a clear steer on aspects of the instrument hardware and analysis that are key or limiting factors governing measurement precision and accuracy. This will enable the reader to make better sense of the multitude of technical details they encounter from Section 2.1 onwards!*

We added such an overview at the end of the introduction:

“In the first part of this paper (section 2) the initial instrumental setup and its subsequent modifications are introduced along with a description of the spectroscopic retrieval technique and the standard operating conditions used. Section 3 discusses and quantifies the residual sensitivities of the measured mole fraction to the thermodynamic properties of the sample, such as temperature and pressure. The issue of inter-species cross sensitivities is discussed as well. This section is extended in Appendix A by a discussion of the thermodynamic conditions in the cell after sample exchange. After characterising the instrument’s sensitivities its response function in the ambient concentration range is investigated in section 4, followed by an exhaustive study on instrument stability and performance on short (hourly to daily) and long (weekly to monthly) time scales in section 5. Section 5 also includes an empirical determination of the necessary calibration frequency. An overall error assessment, including all findings and corrections of the previous sections, is given in section 6. The paper concludes with a discussion of the general applicability of the in situ FTIR analyser for background greenhouse gas monitoring purposes, as well as a recommendation for further improving the instrument performance”.

- *Similarly, it would help if the first paragraph (up to l15 of p3673) in section 6 was moved to section 2, and that the hardware and Standard Operating Conditions for the current Heidelberg setup were all discussed in a dedicated subsection. Once again, it would help the reader to have the key instrument changes required to meet the performance requirements clear at the outset so they can more readily appreciate the relevance of the subsequent results and discussion. A single terminology should be adopted to refer to original and current configurations (the text uses new/modified/Heidelberg/current, and is sometimes difficult for the reader to follow).*

Thanks for this very helpful suggestion: We have included the information from section 6 into the instrument description in sec. 2. In addition we defined a single terminology for the different instrument configurations and used this throughout the MS. This definition of the different instrument configurations summarizes the key instrument changes. By moving the results of the instrument improvement to section 5 we could cancel section 6 entirely.

- *The flow of the presentation of the results would be clearer if the material in Section 2.4 (and figure 2) were moved to an appendix. The discussion of the homogeneity of the temperature distribution within the cell in the introduction to Section 3 is sufficient lead-in to the presentation of the temperature sensitivity results (and the reader can be referred to the appendix for further details). The flush/evacuation sample exchange methods are only mentioned once again themselves in passing in Section 6, and reference there to the Appendix would also suffice.*

We adopted your proposed changes and moved section 2.4 to the Appendix.

Specific comments

Section 2.3

I struggled with this section, which covers a lot of measurement concepts which have not been explicitly introduced (air and tank sample types (calibration procedure requiring reference gases); spectrum acquisition in static and flow modes of operation, flow and evacuation methods sample exchange . . .) and further mixes details specific to original and current configurations of the analyser and its operation. The reasons behind the choice to flow both air and calibration tank samples are never described explicitly (although systematic biases between static and flowing measurements are mentioned in Section 3.5). I believe the unresolved issue of static/flow biases is as important (and potentially related to) the measurement temperature dependencies, and should be discussed more fully.

The whole section was re-organized and is now split into different subsections dealing with the different aspects of the SOC like: Instrument settings; Measurement modes, Measurement intervals and Sample switch over. We introduced and discussed the static/flow bias in more detail and revisited it in the conclusion and outline section as well.

Section 2.4

In addition to the recommendation above to make this section an appendix:

- p3652 l1-2 Reword? e.g. 'Sample measurement and calibration accuracy depend both on the intrinsic instrument performance (SNR) and on the complete exchange of the gas sample in the cell without memory effects.'

Rewording to:

Measurement accuracy depends on instrument precision and on the complete exchange of the gas sample in the cell without memory effects.

- p3653 paragraph 2 Surely some mention of the RTD sensor response time needs to be made here?

In this paragraph we discuss the problem of sample exchange from a more general point of view. Thus the response time of the sensor is only a deficiency of the instrument but nothing which is related to the condition of the sample itself. The effects discussed here are effects on the sample.

Details of the calculation of the Reynolds number for the cell should be given (e.g. in a second appendix).

To better introduce the Reynolds number we have added the following sentence:

"For a tube geometry the Reynolds number can be calculated as: $Re = (um\ d)/\nu$, with um : mean velocity of the gas ($1\ 10^{-3}\ m/s$) in the cell, d : diameter of the cell ($0.15m$) and ν : kinematic viscosity of air ($1.5\ 10^{-5}\ m^2/s$)."

Section 3

I feel the introductory discussion does not distinguish clearly enough between systematic retrieval biases which depend on sample properties (e.g. the retrieved target gas concentration may vary with cell pressure due to errors in spectroscopic parameters describing line broadening) and errors in the measured values of the sample properties (pressure, temperature) and their propagation into the estimated dry air mole fraction. In principle, sensor errors are 'unknowable' random errors which cannot be corrected (systematic biases should be eliminated in hardware).

We included an extended discussion on the different error types.

The systematic retrieval biases can in principle be characterised by sensitivity experiments of the type described here, provided the errors in the measurements of the sample properties themselves are sufficiently small. Alternatively, the empirical sensitivity characterisation can be considered to comprise contributions from the retrieval and sample parameter measurement errors, and this may be useful when considering the temporal stability of a sensitivity correction multiplier: the spectroscopic errors are unchanged, but measurement sensitivity may alter due to changes e.g. in instrument ILS or drifts in the calibration of sample parameter sensors. One would assume the latter (drifts) would small, but for these high-end applications this probably needs to be checked routinely: significant drifts have been found in the pressure sensors (MKS 902 series piezo transducer) used in some prototype FTIR analysers. Given the comments above, the 'Temperature disequilibrium sensitivity' merits careful presentation. The error in the estimated gas temperature can only be considered to be a correctable cross-sensitivity if the error is reproducible i.e. a systematic bias which depends on some other system variable (e.g. flow rate) and the error can be reliably estimated i.e. the error can be characterised, but for some reason, the method used to characterise the error cannot be applied in data acquisition (or at least not retro-actively).

The multiplier dX/dT in the Taylor series expansion for the tracer dry air mole fraction X as a function of temperature error $_T$ can be derived by perturbing the temperature assumed in the Malt retrievals and the associated calculation of the molar concentration of dry air in the cell. The temperature error $_T$ is common to all tracer retrievals, so the

empirically determined tracer sensitivity to cell temperature (in this case) $\frac{\Delta X}{\Delta T}$ should give a consistent estimate of ΔT for all tracers. The signs and magnitudes of the slopes of the curves in Figure 4 are consistent with a common temperature error which is directly proportional to the cell temperature for CO₂, CH₄ and CO but not for N₂O (I did not do the calculation for $\delta^{13}\text{C}$). Given the magnitude of the corrections applied (~ 2 ppm/C for CO₂), this is cause for further thought, particularly as there is no clear mechanism for the inferred temperature error.

We have performed the advised theoretical MALT study to investigate the effect of the density and the introduced spectroscopic error. As you outlined the TDS for CO₂, CO and CH₄ is consistent with a biased (1°C) temperature measurement. For N₂O only 50% of the observed sensitivity is explained by this incorrect temperature measurement. We believe that the biased temperature measurement is due to the large thermal mass of the RTD sensor. The resulting slow response time could also be confirmed by a post-calibration of the sensor after it was replaced. A biased temperature measurement explains as well why the TDS was no longer observed after the RTD was exchanged by the thermocouple.

We have rephrased the entire TDS section and included this discussion in the manuscript. Table1 was extended with the results from the theoretical MALT study.

If I interpret the room temperature bars at the top of Figure 4 correctly (i.e. assuming that the cell temperature range shown for each box is the true range for the given site/instrument deployment), then the data set forms three disjoint clusters (in cell temperature), and in many cases, the temperature dependence of a given cluster is a lot less marked than the impression given by the ensemble. I assume (the authors should state explicitly in the text and figure legend) that the measurements shown are not calibrated.

We included such a statement in the text as well as in the Figure legend. We however don't agree that the temperature dependency is generally less marked in the individual clusters. CO₂, $\delta^{13}\text{C}$, and CH₄ will result in very similar TDS if the individual clusters are evaluated on their own. Only in case of N₂O and CO where additional scatter is observed this remark is valid.

If the mean difference between air and calibration tank sample temperatures (ΔT) is ~ 0 , then would calibration not eliminate the first order site dependent differences? Residual variability of the order of e.g. ± 0.1 ppm full range in CO₂ would remain for each sub-ensemble, but this is arguably a more realistic estimate of the measurement reproducibility, at least in the configuration using the PT100 RTD temperature sensor. If the mean ΔT is not ~ 0 and/or this varies from deployment to deployment, then the situation is more complicated and would suggest to me, as the authors discuss, that the problem is fundamentally related to cell conditioning and/or temperature sensor response post-evacuation.

Yes, calibration accounts for that part of the temperature variability which is covered by the calibration measurements as well. If we assume 48 hourly calibrations the calibration will account for e.g. monthly or seasonally temperature change, but it will not be able to correct diurnal temperature variations. Thus, we think that discussing the effects of calibration in this context is misleading. Our main aim in section 3 is to characterize the instrument response.

Finally, I recommend the discussion of the different TDS with RTD and J-type thermocouple temperature sensors in paragraph 2 of section 6 (l15-24 of p3673) is moved to Section 3 (e.g. create a subsection 3.3 Temperature sensitivity with subsections 3.3.1 and 3.3.2 containing the material currently in sections 3.3 and 3.4 respectively; add the material from section 6 as a third subsection (3.3.3)).

We adopted this and changed the sections

Section 5.1

Had the PT100 RTD temperature sensor been replaced when the short-term stability test was run? If so, the text on p3667 l4-7 appears to contradict the statements in Section 6. Please clarify in the manuscript.

The short term test was run in the Ecotech setup, meaning that the RTD had been replaced by the thermocouple. In section p3667 l4-7 we did not mention explicitly that we corrected for the RTS, it was only stated 'temperature correction'; we explicitly included RTS. In section 6 we discussed that with the thermocouple the TDS was no longer observed. We mentioned explicitly that the RTS remains fairly unchanged.

Section 6

Do you think the conclusions regarding the merits of evacuation and flushed sample exchange depend on the temperature sensor employed in the experiment i.e. do you think a similar conclusion holds for the PT100 RTD sensor or not? It would be worth noting in the manuscript.

We definitely think that this is NOT valid for the RTD temperature sensor. The settling in time for the RTD sensor after the evacuation is far longer. We included a warning that this conclusion might be only valid for the thermocouple and needs further investigations for other temperature sensors:

“This finding depends on the temperature sensor type used. It will most likely not be valid for the slow responding RTD temperature sensor”.

Conclusions

I think the conclusions definitely need to recap. the instrument modifications and operating conditions (flowing calibration etc) that were needed to fulfil the requirements for precision, accuracy and stability. I would add static/flow bias to current outstanding issues: as you say, flowing calibrations require a lot of precious cal gas, and this is a significant drawback particularly for deployments at remote field stations.

We included the instrument modifications and the discussion of the static vs. flow bias in the conclusions.

Technical corrections

- *in situ should not be hyphenated*

Done

- *write out demo-experiment in full i.e. demonstration experiment*

Done

- *does two-daily mean twice daily or every two days? please revise throughout*

Changed to 24- or 48-hourly

Abstract

- *There is a mix of tenses used in the abstract.*

Corrected

- *p3646 l2 suggest reword as 'The FTIR analyser is shown to measure . . . '*

Done

- *p3646 l7-9 suggest reword as 'Residual dependencies between . . . thermodynamic properties of the sample and the cross-sensitivities . . . constituents are quantified and minimised where possible.'*

Extended re-wording changed this section

- *p3646 l11 suggest reword 'investigated by' as 'characterised using'*

Extended re-wording changed this section

Section 1

- *p3647 move paragraph 'Another benefit of optical spectroscopy ...' after the paragraph ending line 24 of that page?*

Done

- *p3647 l21 cut Especially*

Done

- *p3648 l1 'Especially' should be replaced by 'In particular'*

Done

- *p3648 l18 reword 'and their benefits' as 'and the resulting benefits'*

This section was replaced by a extended summary of the paper

Section 2

- *p3648 l23 reword 'Further on' as 'Hereafter'*

Done

- *p3649 l5-6 reword as 'the transfer optics between the interferometer, the multipass cell and the interferometer housing ...'*

Done

- *p3649 l22 Should original be added for clarity in 'For additional stabilization ... in the cell, the original Heidelberg configuration used an add-on ...'?*

Different instrument setups have now a clear and unique naming

- p3651 l17 reword as *'Measurements of ambient air . . . still differ with respect to . . . however.'*

General rewording removed this sentence

- p3651 l21-22 Suggest cutting sentence *'Possible disadvantages . . . later on.'*

We decided to keep the clause for clarity.

Section 3

- p3655 l3-4 The reference to *'current FTIR setup'* here is ambiguous: the sample injection is the same for original and modified versions of the instrument right?

Different instrument setups have now a clear and unique naming

- p3661 l16 reword *'linearity disappeared'* as *'linearity broke down'*

Done

Section 6

- p3675 l28 reword to clarify that CO₂ refers to CO₂ cross-sensitivity

Done