

Response to Reviewer 1

Please find the original comments of reviewer 1 in black and our replies in green.

The manuscript by Schmitt et al. demonstrates the development and deployment of a near-infrared open-path dual-comb spectroscopy system for CO₂ (and CH₄) observation over the city of Heidelberg. The manuscript is well-written and describes the dual-comb system well. While the authors demonstrate the good precision of the system for the concentration of CO₂, the ambiguous position of CH₄ in the manuscript needs to be addressed before this manuscript can be recommended for publication.

Reply: We thank the reviewer for the overall positive feedback concerning the manuscript as a whole and agree with their assessment concerning those parts of the manuscript which address CH₄.

Specific comments:

1. My primary concern in this manuscript is that CH₄ is in some parts of the text presented with a more central position in this work than in other parts. For instance, while the title and abstract do not mention CH₄, the second sentence of the conclusion mentions that the system "allows to retrieve column-averaged dry-air mole fractions of CO₂ and CH₄". The manuscript, however, does not contain any column-averaged dry-air mole fraction of CH₄, though section 3 does discuss the problems in fitting CH₄ in detail. I would suggest either removing the content on CH₄ from the manuscript altogether, if the authors wish to focus solely on CO₂, or (and this would be my personal suggestion) giving CH₄ a more central role in the manuscript, mentioning it in the abstract (and potentially the title).

In the latter case, I would suggest elaborating further on the fits of CH₄: could the fit be improved using different reference spectra, either using observational CH₄ spectra or the older HITRAN08 line list (as mentioned in the HITRAN24 paper p.17, many "remote sensing" groups prefer this version as it gives the most reliable methane fits).

On p.8 of the current manuscript, a spectrum of CH₄ is shown, and the text describes the problems with the fit. However, what are the actual, retrieved concentrations of CH₄ using the current best fits? It would be interesting to learn what the achieved precision is and whether the line list problems only cause a large bias (since the line list is static and its error should therefore also be systematic). Figure 4(a) shows absorption lines with a strength around ~0.1, with residuals of those peaks generally within 0.0025, so one would quickly estimate a bias to be around or lower than 2.5% of the reported value. Is this correct?

Reply: We agree that the situation in the submitted version of the manuscript is not ideal and decided to remove the CH₄ content from it entirely. Since the current situation for the CH₄ line list is rather complex (with some lines receiving updates and some not), a proper discussion is not possible within the scope of this manuscript. As a result, most numbers for CO₂ changed as well, since the retrieval coupled the two gases via the path averaged temperature.

2. The comparison to the co-deployed FTIR yields a positive bias of the DCS instrument of 0.16 ppm CO₂. Since this is a major outcome in this work, further elaboration on the nature of this bias would be appropriate. For example, do the residuals of both instruments for the longest possible averaging time (thus minimal noise) show certain differences in how well absorption lines are fitted? Do both instruments fit the same absorption lines of CO₂ or could the difference be explained by line list discrepancies between the different absorption lines? And is the bias stable over the fluctuations in pressure and temperature, or is the bias a temperature effect? Particularly, an analysis of the bias over varying ambient temperature could be interesting.

Reply: There are no obvious differences in the fit residuals between the instruments, but please keep in mind that the FTIR spectra cover individual lines only with a few samples each. So a one-to-one comparison of the spectra is difficult. Both instruments fit the CO₂ absorption lines at 1.6 μ m and the FTIR instrument makes use of the absorption band at 2 μ m additionally. We made the description in the manuscript clearer. The bias is not a temperature effect and indeed stable over the temperature range from 270 K to 300 K that our data covers (see Fig. R1).

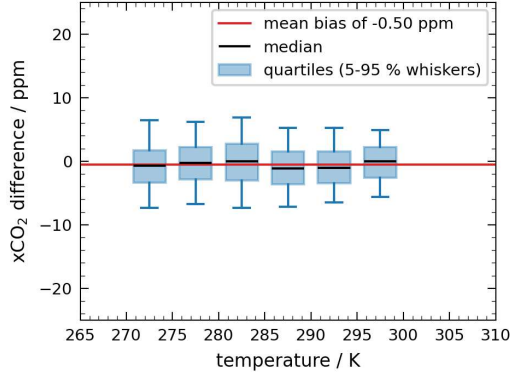


Figure R1: Distribution of the differences between the two instruments as a function of temperature. Binning in intervals of 5 K. Bins below 270 K and above 300 K are excluded due to low statistics.

3. In section 2, the authors describe the data acquisition of the interferograms, which consists of a rudimentary averaging step, centering the maxima of the interferograms. While I understand that the active stabilization in principle enables this approach, I would invite the authors to comment on the level of phase noise their system has and whether this does (or does not) obstruct their averaging without any phase-noise correction.

Reply: According to the commissioning documentation, the single sideband integrated phase noise of the carrier envelope offset (CEO) frequency is about 70 mrad and 76 mrad in a 100 Hz to 2 MHz interval for the two combs respectively. For the beating frequency with the reference laser, the integrated phase noise is 41 mrad in a 100 Hz to 2 MHz interval for both combs. Sporadic, quick re-measurements of these values with a spectrum analyzer resulted in similar values.

Since we measure close to the mutual optical reference, the CEO lock is less relevant and the high frequency phase noise is dominated by the optical beat lock and the relative phase noise of the two combs ($\sqrt{2} \times 41 \text{ mrad} \approx 58 \text{ mrad}$).

Regarding the averaging procedure: both combs share the same CEO frequency, so the relative carrier-envelope phase slip between consecutive interferograms vanishes and the carrier position is fixed with respect to the envelope. Centering on the maximum of the interferogram is therefore unambiguous and simultaneously acts as a first-order carrier-phase correction. Since the correction is applied by shifting whole samples, it is quantized: with approximately 13 samples per carrier period, one sample corresponds to $2\pi/13 \approx 480 \text{ mrad}$. As the interferogram is not an exact integer number of points, the position of the center burst within a sampling interval varies between interferograms, and we conservatively assume it to be uniformly distributed, which yields a residual phase fluctuation of approximately $480 \text{ mrad}/\sqrt{12} \approx 140 \text{ mrad}$. Adding both contributions in quadrature gives approximately 150 mrad, corresponding to a modulation efficiency of about 99%, an order of magnitude below the $\sim 1.4 \text{ rad}$ at which the fringe contrast would drop to $1/e$, which is often given as the coherence limit. While the quoted lock figures are in-loop measurements and thus represent lower bounds, we observed no degradation of fringe contrast or signal-to-noise ratio over the averaging times used in this work. We have added a corresponding discussion to Section 2.

4. In lines 140-143, p.6, the authors describe that fitting pressure from their spectra reduced the CO₂ concentration precision due to the uncertainty in the retrieved pressure. Could the authors provide additional information about this finding, such as a spectral residual plot comparison of both methods or an example of the difference in retrieved concentrations for both methods over time?

Reply: We thank the reviewer for this suggestion and added the following paragraph on the topic at the end of Sect. 4:

Redoing this evaluation for eight hours of data on 30 October 2025 by fitting the atmospheric pressure instead of measuring it results in the following changes: The more degrees of freedom result in a deterioration of the xCO₂ precision by about 16% to 7.13 ppm/s. Further, the fitted pressure results in a 0.28% low bias compared to the fixed pressure evaluation scenario, but also a correlated 0.26% low bias of the retrieved CO₂ column. Together with a 0.10% (0.28 K) low bias of the retrieved temperature, these correlated biases mostly cancel and result in a 0.09% (0.36 ppm) low bias of the retrieved xCO₂. The systematic residuals do not show any significant differences compared to those in

Fig. 3. This encourages the use of fitted pressures, where an accurate pressure sensor is not available or regular recalibration is not possible.

5. How big is the difference between the locally measured temperature and the path-averaged temperature from the spectrum (Line 168, p.8)? It would be interesting if the retrieved temperatures are also shown in the manuscript, similarly to the retrieved concentrations. This would help to interpret the precision of temperature retrieval as well as provide insights into the possible application of retrieving a path-averaged temperature.

Reply: The fitted path-averaged temperatures agree well with locally measured temperatures during the night, but differ substantially during the day (several Kelvin for hot summer days). Most of this is likely not due to actual strong temperature gradients, but dominated by imperfect radiation shielding of the temperature sensor and its inherent position close to a building. The retrieved temperatures have a precision of about 0.5 K for a 1 min time resolution and are the second most relevant contributor to the final xCO₂ uncertainty after the CO₂ columns themselves.

6. The authors provide a comprehensive overview of laser-based open-path spectroscopy, but have a very strong focus on the work developed by NIST. While this group has certainly been instrumental in the development of this technology, I suggest removing the statement "All of these advances originated from NIST, Boulder, Colorado, and collaborating institutions." (line 62, p.3), as readers can interpret this from the references themselves, and "all of these advances" could easily be interpreted as "all advances" which would not do justice to other institutes working on this technique. For example, regarding open-path DCS, the authors could consider the relevant work on QCL-based open-path DCS [1]. Further, while "Open-path Fourier transform infrared (FTIR) [...] [is] fundamentally constrained by the limited spectral radiance of thermal sources" (line 37-40, p.2), this has been demonstrated to be overcome when combining a (laser-based) supercontinuum source with FTIR [2]. And besides the work mentioned on TDLAS, QCLs have also been employed for such laser-based open-path spectroscopy [3].

References:

1. Westberg, J. et al. Urban open-air chemical sensing using a mobile quantum cascade laser dual-comb spectrometer. *APL Photonics* 8, 120803 (2023).
2. Krebbers, R. et al. Ultra-broadband coherent open-path spectroscopy for multi-gas monitoring in wastewater treatment. *Environmental Science and Ecotechnology* 25, 100554 (2025).
3. Rodrigo, P. J. et al. Fast horizontal radial plume mapping of N₂O using open-path absorption spectroscopy with a quantum-cascade laser. *Atmospheric Environment* 120510 (2024).

Reply: We thank the reviewer for the positive feedback on the comprehensiveness of our literature overview and acknowledge that some of our phrasing could be interpreted as overly general. We have revised these passages accordingly. Concerning the suggested extension of the review, we do not agree on all points. We chose to limit the discussion to (a) demonstrations on the km scale and (b) CO₂ and CH₄. Deviating from these constraints would, in our view, quickly turn the section into a general review of open-path techniques, which would then have to include the extensive DOAS literature as well as the many laser spectroscopy methods that each span their own wide wavelength ranges and orders of magnitude in path length.

That said, we added Krebbers et al. (2025) to point to supercontinuum sources, while also stressing that these do not simply allow an established FTIR system with a robust interferometer core to be paired with a new light source, but instead require a different interferometer design and, in particular, a different detection strategy.

7. The statement in the abstract that "The results demonstrate that high-quality, long-term open-path DCS operation is achievable with readily available hardware" (line 13) is not very strong, given that long-term open-path DCS has been demonstrated in many papers by other authors. It feels like a deeper analysis of the results obtained in the work might give a stronger comparison to the existing co-deployed FTIR and other available techniques, resulting in a stronger, more substantial conclusion on the applications enabled by this DCS system.

Reply: We changed the remarked line in the abstract and put more emphasis on the transformative aspect of using only commercially available hardware in contrast to custom-built frequency combs, as is typically the case for previous DCS open-path publications.

Technical corrections:

1. Please check the caption of Figure 8 for language issues. For example, datapoints instead of "data-points", timeseries instead of "time-series", and Gaussian noise instead of "gaussian noise".

Reply: We thank the reviewer for catching this.

2. I would suggest making the processed data (retrieved concentrations, temperatures, etc.) available as supplementary data.

Reply: After consulting with the editor, as per AMT submission guidelines, we will make the 5-min gridded data available as supplementary data with this manuscript.

Response to Reviewer 2

Please find the original comments of reviewer 2 in black and our replies in green.

This paper presents an open-path dual-comb spectroscopy system based on commercial frequency combs. The system was deployed to measure CO₂ (and CH₄) across a 3.1 km round trip path in Heidelberg. The system was operated for 6 months with high uptime. The sensitivity of the system is comparable to or better than previous demonstrations. A particularly interesting/novel aspect of the study is the direct comparison to a co-located open-path FTS running over the same path. The paper shows a remarkably good agreement between the DCS and the FTS, with the DCS providing an order of magnitude better sensitivity. Overall, the paper is well written and suitable for publication in AMT. I do have some minor comments/corrections:

***Reply:** We thank the reviewer for the overall positive review. Please find our point by point replies below.*

1. The Mead et al ref used mid-infrared DCS and detected ethane and propane as well as methane. As such, the citation location is not completely appropriate. It would be better to either cite after Giorgetta 2021 (line 59) or at the end of the oil and gas section as an extension to the mid-infrared.

***Reply:** We agree and changed the placement of the reference accordingly.*

2. What dfr was used? Are the combs locked to have an integer number of points per interferogram?

***Reply:** We thank the reviewer for catching this oversight on our part, since we have not reported the repetition rate difference of about 50 Hz. The combs are not locked to an exact integer number of points per interferogram. We further comment on the resulting implications in our reply to point 7.*

3. What is the transmitted beam diameter? Is it a collimated beam? I could not tell from Fig 2

***Reply:** The transmitting optics are of 2-inch diameter, with the off-axis parabolic mirror collimating the beam at a $1/e^2$ diameter of 12.8 mm. So, yes, the transmitted beam is collimated, but with a Rayleigh range substantially below the kilometer distance. As a result (and as a result of atmospheric turbulence) the received beam is substantially wider. We extended the caption of Fig. 2 to better explain this.*

4. Did you use any active transceiver alignment feedback? Is the transceiver located outside or in a building?

***Reply:** We did not use any active transceiver alignment. The transceiver is located inside, the reflector array outside. We added that to the manuscript.*

5. Did you try a single retroreflector instead of an array? If so, how different was the return power? Do you see any issues in the interferogram or measured spectrum due to the multiple reflectors?

***Reply:** We did not. But we recently switched to an array assembled from gold-coated hollow cube corners, which improves return power significantly. As far as we can tell, there was no loss of modulation efficiency or similar effects in the interferograms as the result of using an array.*

6. What are the integrated phase noise numbers on the comb locks?

***Reply:** According to the commissioning documentation, the single sideband integrated phase noise of the carrier envelope offset frequency is about 70 mrad and 76 mrad in a 100 Hz to 2 MHz interval for the two combs respectively. For the beating frequency with the reference laser, the integrated phase noise is 41 mrad in a 100 Hz to 2 MHz interval for both combs. Sporadic, quick re-measurements of these values with a spectrum analyzer resulted in similar values. Since we measure close to the mutual optical reference, the CEO lock is less relevant and the high frequency phase noise is dominated by the optical beat lock and the relative phase noise of the two combs ($\sqrt{2} \times 41 \text{ mrad} \approx 58 \text{ mrad}$).*

7. When co-adding the interferograms, do you do a phase correction between interferograms? If so, could you provide more information about how this is done? Is it a single phase per interferogram, or a phase and timing correction (e.g., phase and spectral phase slope)? Related to question 2, are the interferograms resampled, or are there already an integer number of points per interferogram?

***Reply:** The interferogram is not an exact integer number of points. We neither resampled the interferograms nor used a full phase correction algorithm. Regarding the averaging procedure: both combs share the same CEO frequency, so the relative carrier–*

envelope phase slip between consecutive interferograms vanishes and the carrier position is fixed with respect to the envelope. Centering on the maximum of the interferogram is therefore unambiguous and simultaneously acts as a first-order carrier-phase correction. Since the correction is applied by shifting whole samples, it is quantized: with approximately 13 samples per carrier period, one sample corresponds to $2\pi/13 \approx 480$ mrad. As the interferogram is not an exact integer number of points, the position of the center burst within a sampling interval varies between interferograms, and we conservatively assume it to be uniformly distributed, which yields a residual phase fluctuation of approximately $480 \text{ mrad}/\sqrt{12} \approx 140$ mrad. Adding the estimated phase noise from the locks in quadrature gives approximately 150 mrad, corresponding to a modulation efficiency of about 99%, an order of magnitude below the ~ 1.4 rad at which the fringe contrast would drop to $1/e$, which is often given as the coherence limit. We observed no degradation of fringe contrast or signal-to-noise ratio over the averaging times used in this work. We have added a corresponding discussion to Section 2.

8. For the spectral retrievals, was the FTS data fit with the same spectral parameters?

Reply: *Where the spectral coverage overlaps, the spectral parameters were the same. Where the FTS instrument had additional spectral information, we used HITRAN2024. We made this more explicitly clear in the manuscript.*

9. I was a bit confused about the uptime numbers. The abstract lists an 85% uptime; however, later a number of 76% is also given. What is the difference between these two numbers?

Reply: *The quoted 76% references the full reported period, including times where the instrument was not available, because it was used for other experiments. The 85% references the latter half of the deployment period, where the frequency combs were dedicated to the open-path measurement full time. We used this number in the abstract, because it better reflects the limitations of the method but understand that this is unnecessarily complicated to explain. For the abstract, we have now switched to the 76% for the total period.*

Response to Reviewer 3

Please find the original comments of reviewer 3 in black and our replies in green.

General Comments:

The authors report details of a measurement with a DCS system used for sensing greenhouse gases in an urban area. They show that DCS can achieve an order of magnitude better precision for CO₂ path-averaged concentration than open-path FTIR at the same location. The application of a turnkey DCS system for precise atmospheric measurements of urban greenhouse gases will be of interest to the EGU community. The article is well-organized, concise, and easy to follow. I recommend the article for publication with only a few minor revisions.

Reply: We thank the reviewer for the positive assessment of our manuscript. Please find our point by point replies below.

Specific Comments:

1. Line 69: The authors cite Eber et al. 2024 as performing “open-path DCS sensing in the ultra-violet”. However, Eber et al. report a DCS measurement from 514 nm to 520 nm, which is in the visible region of the electromagnetic spectrum. This should be corrected.

Reply: We thank the reviewer for catching this mistake. We corrected it in the manuscript.

2. Line 135: It would be helpful if the authors would give an example from the literature of real-time averaging at the hardware level.

Reply: We agree with this statement and now reference Roy et al. (2012) as an early example of real-time averaging at the hardware level and Eber et al. (2025) as a more recent example, which is maybe more accessible for outside adoption.

3. Line 211: The authors note that their system achieves CO₂ precision better than all compared studies other than Waxman et al. (2019). However, from figure 8, the actual allan deviation appears to be worse than Malarich et al. (2025) at 1-hour averaging time. As the authors mention, drift in atmospheric composition drives larger allan deviations beyond 400s, which is why they extrapolate a fit for the first seven points. Since Malarich et al. (2025) was performed at Mauna Loa Observatory, far from sources and sinks, whereas this work was performed in an urban environment, it is reasonable to expect Malarich et al. (2025) to show lower allan deviations at longer averaging times, even if the systems perform similarly. This different measurement context should be made explicit for the reader who may not be familiar with the work of Malarich et al. (2025).

Reply: The reviewer is of course correct in their assessment that the data point of Malarich et al. (2025) resembles a lower scatter/higher precision, but that this is a result of the high averaging times, which are only enabled by the low variability. We think that it is still clear from Fig. 8 (now 7) that this is not a result of instrument performance, but we added a statement addressing the fact.

4. I would like to know whether there is potential for gains in precision with the FTIR system by improving the duty cycle, as is possible with the DCS. Additionally, from looking at figure 5, I am curious about the extent of temporal coverage with the FTIR system and how this compares to the DCS system. Please report the duty cycle of the co-located FTIR instrument and the fractional coverage of 5-minute bins by FTIR.

Reply: The duty cycle of the FTIR instrument can only be estimated from the information the proprietary data format and software provides, but is not explicitly provided natively. An upper bound would be about 95% by dividing the given "scan time" by the total time between two measurements. Calculating it on the basis of the number of points per interferogram and some assumptions about the sampling frequency and decimation yields a lower value of about 70%. Concerning the coverage of 5-minute bins by the FTIR: it is typically very similar to that of the DCS system, since they are both limited by the same visibility constraints. All major differences are down to instrument availability or individual servicing periods of the respective instruments. But for the total time span, the FTIR instrument covers 61% of 5-minute bins, while the DCS covers 76%. But this is mostly dominated by January and February, where the FTIR instrument was not reliably available. In most individual months, the coverage was quite similar. In the past, the FTIR demonstrated coverage of up to 90% under good conditions (Schmitt et al., 2023).