Atmospheric
Measurement
Techniques

Open Access

EGU

Discussions

Interactive comment on "Continuous and high precision atmospheric concentration measurements of COS, CO_2, CO and H_2O using a quantum cascade laser spectrometer (QCLS)" *by* Linda M. J. Kooijmans et al.

Anonymous Referee #2

Critique Summary

In general, I believe that this is a well-constructed experiment of the characterization of the Aerodyne QCLS system for measurement of ambient concentrations of COS, CO2, and CO. I believe that it is appropriate for publication in Atmos. Meas. Tech. after substantial revisions. My major criticisms are related the presentation of uncertainties and how they relate to different types of analyses that might be performed with the data set collected at Lutjewad. Presumably, the authors have written this manuscript with future analyses in mind that focus on interpreting ambient COS, CO, and CO2

data. The particular type of analysis will determine which types of uncertainty should be considered. I suggest revising the manuscript in a way that makes it clear in the Introduction what types of analyses are intended (or are possible) for this data set, while also discussing the theoretical requirements for uncertainty in these cases. The authors hint at two ways in which the data might be used – interpreting vertical gradients in COS at Lutjewad and interpreting horizontal gradients between a network of stations reporting COS (as in Figure 11). The latter use of data requires a much more stringent evaluation of accuracy since it requires the data from all stations to be on the same scale. The former, has a seemingly less stringent need for accuracy, but instead is more dependent on short-term precision in order to quantify what might be small differences in concentration from one height to the next. This is my impression, anyways, and I suggest that the authors explore this idea and report on whether there should be two different estimates of total uncertainty depending on the type of analysis (i.e. vertical profiles vs station-to-station differences).

For this review, I first provide a general critique of the authors' presentation of errors in the QCLS measurement, including a discussion of the link between uncertainty type and the intended analysis, as suggested above. Then, I provide a list of specific comments and edits for the authors to address.

General Critique

On the evaluation of target cylinders of COS, CO, and CO2 for accuracy against the NOAA/WMO scale (Figure 9 and Sec. 3.1), I would expect that the mean bias should be lower than the uncertainty listed in Table 5 as "Transfer to calibration standards" (2.8 ppt for COS). Is this correct? Or should it be compared to the quadrature sum of "Transfer to cal standards" and "Calibration using calibration standards" (4.0 ppt for COS)? In any case, this comparison needs to be made in order to evaluate whether the uncertainties listed in Table 5 (and contributing to the total uncertainties) properly envelope the bias determined from this experiment. Using COS as an example, the mean bias is -3.3 ppt and -1.7 ppt for the two cylinders (single bias correction). The standard error of the

mean should be given, as well. Even though the mean is within 1 standard deviation of 0, with so many measurements made, the mean bias should be very well known and will likely be significantly different from 0 (the standard error will more accurately express how well the mean is known than the standard deviation). The next question becomes, is this mean bias (+/- the standard error) reflected in the uncertainties listed in Table 5. Additionally, please also give the absolute concentrations for the cylinders used in this experiment. It is relevant to know whether there is an increased bias for the target cylinders at concentrations outside of the range of the NOAA/ESRL standards. I use COS as an example here, but this comment also applies to CO and CO2.

Next, I think it is important to distinguish between short-term precision and repeatability and provide some discussion about where each measure of uncertainty should be used. As written, the authors do not sufficiently make this distinction. Table 5 lists the repeatability uncertainty as the 1-second standard deviation from the histograms in Figure 8. I would not consider this value a measure of repeatability, rather, it should be the standard deviation of the mean of the repeat measurements on the target cylinders in Figure 9. I would expect the repeatability to be somewhat higher than the short-term precision because typically there is some amount of instrumental drift. The repeatability for COS is 6.6 ppt and 4.9 ppt for the single bias correction case and 5.3 ppt for the hourly response curve case. All three measures of repeatability are somewhat higher than the short-term precision, and the total uncertainty (Table 5) should probably reflect these higher values.

When the repeatability (rather than the precision) is incorporated into the total measurement uncertainty, the total uncertainty has relevance for interpreting diurnal, synoptic, seasonal, and longer-term variability in concentrations. It is also relevant when comparing the Lutjewad station measurements to measurements at other sites and with other measurements that are referenced to the NOAA/WMO scale. Therefore, an analysis of data in Figure 11 should have a total uncertainty attached that includes repeatability rather than short-term precision. On the other hand, if the goal is to inter-

C3

pret differences between one height and another in the vertical sampling arrangement at Lutjewad, this is where the short-term precision is relevant. In this situation, one is looking for small differences between two (essentially) simultaneous measurements. Additionally, since this type of analysis relies on a comparison of measurements from the same instrument, I would expect that there should be less reliance on the need to transfer the measurements to the NOAA/WMO scale (and the total uncertainty should therefore be lower). Is this correct? If it has been determined that a background offset correction is more important that correcting for a variable instrument response curve, then the near-simultaneous measurements at both inlet heights will be affected by the same offset correction, and, as a result, uncertainty in that offset correction will cancel when estimating the difference between measurements.

In any case, it is important to distinguish between uncertainties that are associated with a comparison of multiple measurement sites vs those associated with inlet height differences. Additionally, the water vapor correction uncertainties may also cancel out unless there is a very strong vertical H2O gradient. While this might be challenging to determine, given the potential for sampling line artifacts, some consideration of this would be worthwhile.

Finally, I question what value there is in reporting the 1-second precision. I would think that unless there is a need to understand high frequency variability in the atmosphere (such as for eddy covariance fluxes), the precision should be reported at a more relevant dwell time. Figure 10, for example, shows hourly averages. Daily averages are used in Figure 11. The 1-second precision has a negligible contribution to uncertainty in the mean on these time scales. For the vertical profile measurements, each height has a 8 minute dwell time. Variability at time-scales less than 8 minutes, therefore, is not important for determining inlet height differences. Even at the 95th percentile of standard deviations for the entire data set in Figure 8 (13.2 ppt for COS), the precision for the 8-minute averages is still very small (likely < 1ppt). The point is, that this short-term precision determines the magnitude of inlet height differences that can be

C4

considered significant. The authors report that differences between heights 40 m and 60 m are, on average, 0.7 ppt +/- 9.7 ppt. It seems, therefore, that the mean difference between the two heights is probably not significant, but with a 9.7 ppt standard deviation, there are likely times when the differences could be interpreted as real (with high confidence) and potentially relatable to the quantification of a COS flux. I think a short discussion of these points would add a lot of value to this manuscript.

Specific comments:

Page 2, lines 31-33: Regarding the need to constrain COS to 1.5 to 4.0 ppt to constrain CO_2 fluxes matching a 1 ppm change. It is my understanding that LRU (leaf-scale relative uptake) scales with the ratio of CO_2/COS mole fractions. In this case if CO_2 is ∼400ppm and COS is 500ppt, then a 1 ppm draw-down in CO_2 would scale to $1.9 - 5.0$ ppt (or $0.4 - 1\%$) in COS. This reflects just a small change from the $0.3 - 4.0$ ppt range quoted, but should be noted, nonetheless.

Page 3, 1st paragraph: Is this paragraph a justification for making vertical gradient measurements of COS in order to estimate fluxes rather than via the eddy covariance method? If so, state this explicitly. It would also be relevant to discuss a few details about flux-gradient theory and the pros and cons of estimating fluxes by this method. For example, doesn't this approach also require a flux measurement of at least one species for which you also have a vertical profile for (e.g. water vapor or CO_2)? I understand that this is a paper focused on instrument characterization, as is entirely appropriate for AMT, but it would be good to also provide some context for the long-term vision of this measurement set-up and how COS (and therefore CO_2) fluxes will ultimately be estimated.

Page 5: Around this section, I started having a hard time keeping the cylinder types straight. I think it would be useful to have a Table that lists each cylinder as either a NOAA/ESRL standard, a reference standard, or a target standard, says which cylinders were brought to the field, and gives a list of what species are characterized in each.

Also relevant for the COS cylinders would be the cylinder material (coated or uncoated) for each.

Further, the COS concentrations in the target/reference cylinders shown in Figure 7 do not match any of the concentrations listed in Table 2. Are these a separate set of cylinders? Which ones are used in Figure 9 as the target cylinders and which ones are used to derive the hourly response curves? Which ones were observed to have some drift to them?

Page 6, line 29: Do you mean to say "uncorrelated with instrument parameters, such as temperature"?

Page 8, lines 16-17 (and Table 5): Could you provide more flexible uncertainties for the water vapor correction for COS, CO_2, and CO, instead of at a single water vapor concentration and single analyte concentration? The most straightforward way to do this, it seems, would be to give the uncertainty on the slope. For example, COS changes by 2.9% (+/-??) per % H_2O. Related to this, a value of 3.5 ppt is listed in Table 5 as the water vapor correction uncertainty for a range of COS concentrations and H_2O concentrations (up to 2.1%). I would expect the uncertainty to scale with H_2O, so if the uncertainty at 1.5% is 3.5 ppt, the uncertainty at 2.1% would be 4.9 ppt. Is this accurate? If not, then why? If so, this should be acknowledged someway in Table 5 and also reflected in the total uncertainty.

Page 11-12, Sec. 2.6: I was confused by the goals of the flask analyses in this study. I was expecting that they would be used to compare the QCLS with the GC-MS measurements at NOAA; however, the discussion here neglects to actually discuss this comparison (although it is shown in Figure 6). Instead, the discussion seems to center on how well the flask measurements via QCLS reproduce the cylinder concentrations that are estimated from in situ measurements. This reduces, essentially, to a test of whether any bias is added with the flask filling procedure or from storage effects. I can understand how this would be important to confirm ahead of a comparison of GC-MS

measurements (which requires flasks) with the QCLS; however, no such comparison is presented in the text. From Figure 6 it appears as though the QCLS flasks are slightly elevated with respect to the GC-MS measurements. Is this significant? Or is it within expectations based on uncertainties in the transfer of the NOAA scale?

Page 13, lines 4-6: Here it is stated that a few uncoated cylinders drifted at a rate of 2-3 ppt per month. Which cylinders are these? How was the drift determined? In previous sections comments have been made indicating that cylinder drift was not found to impact certain experiments. Given that COS drift in cylinders appears to be a persistent concern, I think the authors should consider adding a new section somewhere that details how the presence or absence of drift was monitored and which cylinders drifted over what time periods (and how that relates to various field and laboratory experiments).

Pages 14-15, Sec, 3.2: I suggest that the authors consider putting this section last since it is not related to characterization of the instrument performance. As written, it falls in between two sub-sections that relate to instrument characterization (Sec 3.1 Precision and Accuracy and Sec. 3.3 Measurement Comparison).

Pages 15-16, Sec. 3.3: What is the point of the flask vs in situ comparison for COS in this section? I was expecting that there would be a comparison between the QCLS and GC-MS flask measurements, given that the section describes an instrument intercomparison for CO and CO2 measurements. What value does the comparison between in situ and flasks on the same instrument add for this site? Is there a longer-term record of COS from flasks at this site that one might want to harmonize with the in situ measurements going forward? Is there some other value to having both flask and in situ measurements at the same site? Is the intent to send flasks to other sites (to be measured by QCLS) for comparison to Lutjewad? Or is it simply a check on the water vapor correction? If so it should be discussed along with the laboratory experiments on the water vapor correction.

Page 16, line 6: "…against -0.02 ppm…" Should this value be -0.12 ppm?

Tables 2 and 5: It appears as though the average of the uncertainties of the 4 cylinders in Table 2 (scenario C) is how the transfer uncertainty is estimated in Table 5. If so, please state this somewhere or, perhaps, include another column in Table 2 that details the average uncertainty for each of the three scenarios. This would help clarify where some of the numbers in Table 5 come from.

Table 4: Include uncertainties on the slopes of the correction curves.

Table 5: Might it be more appropriate to rename "Calibration using calibration standards" as "Calibration of the Instrument Response"? This might clarify the meaning of this uncertainty.

Figures 3 and 4: Consider calculating the mean and standard error of each grouping of data (in between background measurements) during the water vapor experiments (in Figure 3) and deriving the linear regressing using these means, rather than the real-time data (in Figure 4). This would certainly clean up Figure 4 (for CO and COS, anyway), but it also might provide more robust fits to the dry/wet ratios.

Figure 5: This diagram added to some confusion about which cylinders were brought to the field site, since only two calibration cylinders are shown. Considering adding some representation of the actual number of cylinders to this diagram, or specifying in the caption.

Figure 6: Consider showing the mean deviation (across all 4 cylinders) for the QCLS and for the GC-MS measurements (with and without outliers included).

Figure 12: I think that the COS figure should be a separate figure, since the comparison figure on the right-hand side is not showing instrument comparability of COS, as it is for CO, CO2, and H2O. It is a bit misleading, otherwise.