

We greatly thank the reviewers for their work. Their comments helped us to improve the structure and presentation of our manuscript considerably. We first make a few general comments regarding the major changes in the manuscript, after that we provide a point by point reply to the reviewer's comments.

Both reviewers commented on the fact that we did not present a more detailed analysis on interpretation of the COS signals (e.g. diurnal variations, land-sea differences and differences between heights). We agree that this would make an interesting study but we think it will be more appropriate to publish this in a separate manuscript. In the introduction of the revised manuscript we state more explicitly that the aim of this research is to test the performance of the instrument and in the results section we discuss the suitability of the QCLS for different potential purposes based on the uncertainty analysis results. The latter also relates to a comment of Reviewer 2 who raised the point that different types of analyses have different requirements for accuracy and precision. E.g. comparing heights at one site with the same instrumentation doesn't require high accuracy. In the revised version we discuss the required precision for the flux-gradient method in the introduction section and we added a discussion about the relevance of different uncertainty contributions for different types of analyses.

Both reviewers commented that the measurement comparison for COS in Figure 12 was misleading, as this was not an instrument comparison, while it was for CO₂, CO and H₂O. To solve this we moved the flask comparison between the QCLS and GC-MS from section 2.6 into the measurement comparison section. Also the different analysis types are better introduced to make the difference between the comparisons clear. Besides that, we present the flask – in situ comparison for COS (top plot of Fig. 12) in a separate figure to avoid confusion.

Furthermore, we got multiple comments regarding the timescale of the presented precisions as not only the one-second precision, but also precisions on longer timescales are relevant to reflect the influence of instrumental drift. To address this, we added a discussion in section 3.1 about two-minute precisions as obtained from Fig. 8.

In the following we will provide a point by point reply to the reviewers' comments. Our replies are written in blue, and corrections in the manuscript are written in green. References to figures, sections and pages in our reply refer to the original submitted manuscript, unless stated otherwise. In the revised manuscript changes are marked in red.

Anonymous Referee #1

In their study, the authors have tested a recent model of the QCL Mini Monitor spectrometer (QCLS) from Aerodyne Research Inc. for its suitability to obtain continuous and high precision atmospheric measurements of COS, CO₂ and CO. They evaluated the performances of this instrument in the laboratory then in the field at the Lutjewad monitoring station in the Netherlands. The ultimate goal is to provide high-quality observations of COS needed to understand the seasonality, the long-term trend and the spatial distribution of sources and sinks of this trace gas because existing monitoring networks of CO₂ and CO make use of other instruments (e.g. cavity ring-down spectroscopy (CRDS) analyzers).

General comments.

Although the QCLS was set up for continuous in situ measurements at different heights at the tall tower of the Lutjewad monitoring station, the authors do not demonstrate the suitability of the QCLS analyzer to perform profile measurements because data recorded at 7m height are not reported. I think that this question cannot remain open after having put much effort into assessing the performances of the instrument. If the authors consider that their data are not suitable for publication, they could instead make use of the hourly measurements of calibration standards conducted while the instrument was in the field in March-April 2015 to simulate the suitability of the QCLS analyzer to perform profile measurements (see below). The continuous COS, CO₂ and CO observations from Lutjewad are not even used to determine trace gas concentrations characteristic of air masses having either dominant continental or marine influence. Examples of diurnal variations could have been provided. Is there a seasonal change in the amplitude of the diurnal variations recorded by the instrument? Instead, the authors focus on COS seasonal variations reconstructed from incomplete records and compare QCLS daily averages with flask pair means from individual daytime sampling events (NOAA/ESRL data, see Fig. 11). If COS undergoes strong diurnal variations at Lutjewad, only daytime data should be compared with NOAA/ESRL data. My overall feeling is that the potential of the QCLS analyzer is not demonstrated in the field.

We agree with the reviewer that not all the potential analysis of these data is shown (daily cycle / profile etc.). We also think that this sort of analysis would be nice to have; however, the aim of this manuscript is to evaluate and improve the performance of the instrument, and not to focus on the interpretation of the COS observations. At the same time we show the measurements obtained at the Lutjewad site, as the reviewer pointed out, to demonstrate the potential of the QCLS analyzer in the field. Accordingly, we show the measurement record and a comparison across sites, and added a discussion on the observed gradients and the suitability of the QCLS to capture these gradients. Our plan is not to present further interpretation/analysis of the observations beyond these in this manuscript, but to publish this in a separate manuscript where we can present an in-depth interpretation. In the revised version of the manuscript we state explicitly in the introduction section what the goal is:

In this manuscript we aim to evaluate and improve the performance of the instrument. [...] Furthermore, we evaluate the total uncertainty of the measurements by

combining the uncertainties of scale transfer, water vapor corrections and the measurement precision. Based on the precision and accuracy that we derive from these experiments, we discuss the suitability of COS measurements on this instrument for different purposes; that is, for interpreting profile measurements and comparing concentrations across sites.

Instead of providing further interpretations of the COS observations in the field we added a discussion in section 3.1 regarding the suitability of the QCLS for different potential applications, as was also suggested by Reviewer 2.

We observed that COS concentrations between 40 and 7 m differ on average by 1.7 ± 5.3 ppt during daytime and 11.4 ± 33 ppt during nighttime. Given that the gradients should be larger than the uncertainties mentioned above, the daytime gradients are too small to be able to use the flux-gradient method. The suitability of the flux-gradient method is further dependent on the choice of the measurement height (gradients are larger closer to the surface), the size of the fluxes at a given site and the turbulence conditions. Even though the use of the flux-gradient method is limited by the measurement uncertainty that we found here, COS profile measurements from this instrument can still be useful to derive storage fluxes, as nighttime gradients are typically larger.

Furthermore, we agree with your point that we should only compare the Lutjewad daytime data with the NOAA/ESRL data and have modified Figure 11 such that the Lutjewad data only include daytime concentrations. The data presented in the new version of Figure 11 do not differ significantly from that in the first version and therefore the main message remains the same.

Thanks to rigorous testing in the laboratory, the authors clearly demonstrate that an appropriate direct water correction for COS is not possible with the combination of the standard spectral fit and the TDLWINTTEL software correction on. A potential solution to this is to modify the standard fit of the spectral bands (hardly achievable by a non-specialist) or to apply homemade water corrections with the TDLWINTTEL software correction on or off. One of the correction strategies (i.e., homemade water corrections with the TDLWINTTEL software correction off) was tested in the field against dry air flask sample QCLS measurements for COS. Data reported in Fig. 12 suggest that COS mole fractions are not well corrected for water interference yet. The upper panel (COS one) could benefit from having H₂O displayed.

The difference between the flasks and in situ measurements of COS gave -3.5 ± 8.6 ppt where the 8.6 ppt standard deviation indeed suggests large variations between the different comparisons. We did a more thorough analysis on these flask comparisons and found two errors in the flask – in situ comparison: the first was related to the transfer of scales to the flask data and the other is that we first assumed that the flask and in situ sampling have the same time delay, while in reality this delay was in the order of 10 minutes. This was corrected for in the CRDS-QCLS comparison, but not for the comparison with flasks. With these corrections included we find a difference (flask – in situ) of -9.7 ± 4.6 ppt, where the flasks

are larger than the in situ measurements. The difference is now more consistent between the different flasks as the standard deviation is lowered to 4.6 ppt, but the mean difference is larger. The average difference of -9.7 ppt could be caused by the uncertainty of the flask and the in situ measurements (where the latter includes the uncertainty of the water vapor correction), the sampling bias of both in situ and flask measurements, as well as the storage effect of air in the flasks. As multiple factors are influencing the comparison, we cannot simply assign the difference to an error in the water vapor correction. Furthermore, we found that after applying the water correction to the in situ data, the differences between flasks and in situ measurements does not show a correlation with H₂O, while without water correction there is a correlation. We added the following in Section 3.2 (revised version) to discuss this:

The average difference of -9.7 ppt could be caused by the uncertainty of the flask and the in situ measurements (where the latter includes the uncertainty of the water vapor correction), the sampling bias of both in situ and flask measurements, as well as the storage effect of air in the flasks. In Table 5 we quantified the uncertainty related to the transfer of the scale to working standards and measurements, the water correction, and the measurement precision. Combining these uncertainties gives an uncertainty of 5.8 ppt for the flasks and 6.5 ppt for the in situ measurements (with the uncertainty of the water vapor correction included). The combined uncertainty of both methods is 8.7 ppt, which is only slightly smaller than the -9.7 ppt difference that we found. We were not able to test the effect of biases resulting from the sampling of flasks. We find that after applying the water correction to the in situ data, the flask – in situ difference does not show a correlation with H₂O ($R^2 = 0.01$; slope = 3.9 ppt COS per % H₂O), whereas without water vapor correction there is a correlation with H₂O ($R^2 = 0.36$; slope = 26.3 ppt COS per % H₂O), which demonstrates that our laboratory derived water vapor correction properly corrects for the effect of water vapor.

The authors provide evidence that the baseline measurements typically done several times per hour by former users of QCLS analyzers (background measurements are determined with high purity nitrogen) are not recommended for continuous and high precision atmospheric measurements. They developed an optimal background and calibration strategy to ensure accurate field measurements and put much effort into stabilizing the temperature of the QCLS analyzer. Unfortunately, I found little evidence in the text and the figures that improving the temperature stability by actively cooling the electronics section and putting the analyzer in an enclosed box improved the precision and accuracy of field measurements. Plots showing the temperature dependency of COS, CO₂ and CO concentrations are missing. The histograms of standard deviations shown in Fig. 8 are not convincing because other parameters than temperature could have affected the noise of the instrument from August 2014 until February 2015 (instabilities of the laser for example that can be responsible for increases in SD). I noticed periods of higher stability before February 2015 (for example in September 2014) than after late March 2015 when the instrument was properly insulated.

Regarding showing data of the temperature dependence of concentrations we did look for a ppt/degree factor for the particular experiment that is described. We did not find a stable correlation between temperature changes and concentration drift for the test in section 2.2.3,

likely because the changes only reach up to 0.06 °C in 30 minutes. From later field work we did find that concentrations drift together with temperature on the order of 100 ppt (COS) per degree change in the electronics temperature. We added this information in Section 2.2:

Hourly measurements of reference standards in the field (during a later field campaign in Hyytiälä, Finland, not shown here) showed that concentrations drift on the order of 100 ppt (COS), 2 ppm (CO₂) and 10 ppb (CO) per degree change of the electronics temperature in a few hours.

It is right that the temperature stability did not improve the instrument precision and we do mention this in line 29-31 of Page 12 and in the conclusion section (line 22-26, page 17). Still, further stabilization of the temperature does improve the setup in the sense that the use of cylinder measurements to correct for temperature-related drift is reduced, which we mentioned in line 19-20, page 7. In the first version of our manuscript we did mention already that also the alignment can likely affect the instrument precision. In the revised version of the manuscript we marked moments where the mirror alignment was changed, which corresponds with moments that the size of the precision changed and indicates that the mirror alignment affects the instrument precision. Besides that we marked a period where the temperature of the room temperature was characterized by rapid changes caused by an air conditioner, which also corresponds with larger standard deviations. We added the following discussion in section 3.1 about this:

For the period marked in red the room temperature was characterized by rapid changes caused by an air conditioner. This was also a period where the instrument precision was adversely affected, indicating that temperature stability influences the instrument precision. [...] What we do see is that the instrument precision changed every time that the mirror alignment was changed, indicated by blue vertical lines. During periods that the mirror alignment was not changed (January 2015 and April 2015) the instrument precision was stable. These results show that the instrument precision can largely be affected by the mirror alignment.

The QCLS measurements of COS were not compared with that of other instrumentation except when four pairs of flasks filled with dry calibration gas were sent to NOAA/ESRL headquarters and analyzed there with GC/MS. Data shown in the top panel of Fig. 12 are misleading as the flasks shown are not analyzed with a different instrument.

See also our general reply.

It is right that Fig. 12 does not show an instrument comparison for COS, that is also why we called section 3.3 “Measurement comparison” and not “instrument comparison. We decided to move the flask comparison between the QCLS and the GC-MS to the measurement comparison section (3.2 in the revised manuscript) and separated this section over three paragraphs:

- 1. Comparison of COS measurements by QCLS and GC-MS;*
- 2. Comparison of flask and in situ COS measurements by QCLS;*
- 3. Comparison of CO₂, CO and H₂O measurements by QCLS and CRDS.*

The three different measurement comparisons are now introduced so that it is clear that the comparison of flask vs. in situ measurements by the QCLS is not an instrument comparison.

Moreover, we present the flask – in situ comparison for COS (top plot of Fig. 12) in a separate figure to avoid confusion.

Because this study is the first in-depth appraisal of the performances of the most recent model of the QCL Mini Monitor spectrometer (QCLS) from Aerodyne Research Inc., it should be published but after major revision.

Improvements are necessary to demonstrate that this instrument is or is not suitable to obtain continuous and high precision atmospheric measurements of COS, CO₂ and CO.

See also our general reply and the corrections described in our first comment.

Since two representatives of the Aerodyne company are in the list of co-authors it would be interesting to know whether the TDLWINTTEL software will offer in the future the possibility to deselect the standard fit and use the new split fit (no more need to perform homemade water corrections). If the instrument drift between two backgrounds can be properly corrected using a reference cylinder, can the TDLWINTTEL software be updated to automatically correct for drift?

Newly delivered instruments will have the software fitting parameters set as the split fit. For instruments that were delivered with the standard fit the Aerodyne representatives are willing to help setting up the split fit using remote access to the instrument software.

Online automatic correction based on a reference cylinder is not preferred, as the cylinder value at the beginning and end of the measurement period is needed to do such corrections. Measurements within the period can't be corrected yet as the direction and size of the drift is not known while the calculations are being done.

Specific comments.

Abstract.

Page 1 - line 15. Hope that the authors will be able to demonstrate the suitability of the QCLS analyzer to perform profile measurements too.

See also our general reply and the corrections described in our first comment.

Page 1 - line 27. The comparison of in situ QCLS measurements with measurements from flasks is not adequate for COS. See also page 9 – line 24, page 12 – line 7 and page 16 – line 9. Data reported in Fig. 12 suggest that COS mole fractions are not well corrected for water interference yet.

See the reply to the second general comment.

Introduction.

Page 2 - line 11. The past study of Belviso et al. (2013) relied on in situ samples analyzed on-line (not flasks) with GC/MS. I am aware of other on-line measurements of COS with GC/MS using the MEDUSA instrument (AGAGE network).

We thank the reviewer for pointing this out. We removed the Belviso et al. (2013) reference in this particular sentence and added a sentence about in situ measurements with the GC-MS technique and with a reference to Belviso et al. (2013):

Although the GC-MS technique can be used for in situ measurements (Miller et al., 2008; Belviso et al., 2013), this technique does not typically allow for high frequency measurements of 1 to 10 Hz.

Page 2 - line 30. Please clarify your calculations. How can you jump from LRU ratios (the ratio of the deposition velocities of COS and CO₂) to measurement precisions of COS better than 0.3-0.8%?

*It would have been clearer if the precision would be stated as 1.5-4.0 ppt COS (i.e. 0.3-0.8 %), where 0.3 is $1.5/500*100$. In fact “0.3 – 0.8 %” is no critical information so we decided to leave these numbers out to avoid confusion. Reviewer two also corrected us in the sense that the LRU scales with the CO₂/COS mole fractions and that this actually translates to a COS measurement precision of 1.9 – 5.0 ppt. Taking these comments together we have corrected the sentence such that it now reads as:*

*If we were to infer the gross fluxes from chamber measurements with CO₂ measurements better than 1 ppm, then, given the leaf-scale relative uptake ratio (LRU) of COS/CO₂ 1.5 – 4.0 (Stimler et al., 2010b; Seibt et al., 2010; Berkelhammer et al., 2014), our goal would be to have measurement precisions of COS better than **1.9 – 5.0 ppt for COS (calculated from LRU and scaling with the ratio of COS/CO₂ mole fractions gives e.g. $1.5*(500/400) = 1.9$ at the ambient level of 500 ppt COS and 400 ppm CO₂).***

Page 3 - line 12. First clear indication that you are interested in profile measurements. See also page 10 – line 8, Fig. 5 and the legend of the figure. Again see page 10 – line 17. The goal was to provide high-quality observations of COS needed to understand the vertical distribution of this gas at Lutjewad but may be the authors consider that their data are not suitable for publication.

See also our general reply and the corrections described in our first comment.

Experimental setup.

Page 3 - line 13. Not sure that the first prototype of QCLS required little operator attention. I think that the recent one still requires operator attention and manual post processing of data.

The question if the QCLS requires much operator attention is indeed a point of discussion. We agree that the first prototype in Stimler et al. (2010a) may have required more operator attention than the analyzers that are nowadays available, but in fact it was stated in Stimler et al. (2010a) that this analyzer required little operator attention. To avoid discussion about this point we removed the part “.. and requires little operator attention (Stimler et al., 2010a).”

Stimler et al. 2010a is in fact Stimler et al. (2009). See also page 4 line 17.

We are not sure about that. The year of publication in the pdf version of the manuscript that is available at the Wiley Online Library states 2010. I suspect that it was first online in 2009 and got actually published in 2010. We stick to referring to Stimler et al. 2010a.

Page 3 - line 17. Unfortunately not for COS.

In fact we do compare with GC-MS measurements using flasks. Based on the comments from both reviewers we decided to better introduce the goal of the flask measurements and stress that these were also meant as a comparison between instruments. The following lines were therefore added in the introduction:

Besides in situ measurements, flask or canister measurements can be a valuable tool for providing information about ambient concentrations of COS as well. For example, flask measurements were used before when constructing an historical record from firm air (Montzka et al., 2004), during field campaigns (White et al., 2010; Blonquist et al., 2011) and for long-term monitoring (Montzka et al., 2007). [...] In addition to the experimental setup for continuous in situ measurements we developed a setup to analyze flasks, which we used to make a comparison with GC-MS measurements of flasks and to assess the laboratory derived correction for water vapor interference.

Page 3 - line 27. Please use the same terms throughout the manuscript. Are primary standard, secondary standard and target gas more appropriate terms? In general, it would be valuable to know how long was the duration of injections of standards.

We used the terms NOAA/ESRL standards (which are in fact calibrated standards against secondary standards at NOAA/ESRL) and calibration standards, which can be either reference or target standards (see line 26-30, page 4 and line 21-23, page 5 in the first version of the manuscript). To avoid confusion we changed the terminology of the standards as follows:

- *"NOAA/ESRL calibration standards" for the ones that were calibrated by NOAA/ESRL.*
- *"working standards" for the ones used in the field (which we called "calibration standards" before) and which are calibrated against the NOAA/ESRL calibration standards in our lab. These working standards can either be reference or target standards depending on how they are used.*

We hope that some confusion about the different cylinders is also taken away now we trace the cylinders throughout the manuscript (see also our reply to the comment on Table 1).

We added a sentence in the caption of Fig. 2 stating the length of the cylinder measurements.

Page 7 - lines 8-10. The statement "The frequency of reference cylinder: : ." is not sustained with data. Show plots of COS vs T for example as in Berkelhammer et al. (2014). See also the discussion about temperature stability page 10 – line 25.

See also our reply to an earlier comment; we added a sentence in Section 2.2 stating the typical concentration changes per degree that the electronics temperature changed.

Page 8 - line 13. Here you mention the existence of instabilities of the QCLS. It is important to know whether instabilities of the QCLS occurred when the instrument was deployed in the field (see Fig. 8). See also page 12 – line 31. Factors such as alignment and laser instabilities influence the precision as well.

Yes, the instabilities occurred both under laboratory and field conditions. For field conditions it is depicted in Fig. 8 during the periods when the instrument was less stable. We added a sentence in section 2.3 saying that we further discuss the instrument precision and its variation over time in section 3.1.

We will further discuss the precision of the measurements and its variation over time in Section 3.1.

Result and discussion.

Page 13 – line 1. How was the overall uncertainty calculated? Please clarify the method.

The overall uncertainty is calculated as the square root of the sum of the squared individual uncertainties. In the revised version of the manuscript we added this information in that particular paragraph:

The overall uncertainty is calculated as the quadrature sum (square root of the sum of squares) of the individual uncertainties.

Conclusions.

This is a nice summary but what are the main conclusions of the study? Is the instrument suitable for tower profile measurements? What perspectives can you draw?

See also our general reply. In the revised version we added a discussion about the suitability of the QCLS to do different sort of analysis (differences across sites and interpreting profile measurements). Also in the conclusion section we added a sentence about this:

We discussed the relevance of different uncertainty contributions for different types of analyses (e.g. differences across sites and interpreting profile measurements). Our setup with the QCLS provides sufficient accuracy and precision to detect gradients larger than 5.2 ppt for COS, 0.11 ppm for CO₂ and 1.4 ppb for CO. For the current setup in Lutjewad, the daytime gradients were too small compared to the measurement uncertainty to be able to use the flux-gradient method.

Table 1. Three cylinders are mentioned in the text but results of two cylinders are reported in Table 1. I suggest to trace the cylinders throughout the whole manuscript.

This was also raised by Reviewer 2 and is a good suggestion. We added the following sentence in section 2.2.2.

To trace the three working standards that we used in the field we refer to these with numbers #1, #2 and #3 throughout the manuscript. Other working standards that were used in this manuscript but not in the field we refer to as “A”, “B”, “C” and “D”.

We made this consistent throughout the manuscript: Table 1, 2 and Fig. 5, 6, 8, 9 (numbers referring to the revised manuscript).

Table 5. It is impossible to reconstruct the overall uncertainty from data gathered in Table 5.

The overall uncertainty is calculated as the square root of the sum of the squared individual uncertainties:

$$\sqrt{2.1^2 + 2.8^2 + 2.8^2 + 3.5^2 + 4.3^2} = 7.1$$

In the revised version:

$$\sqrt{2.1^2 + 2.8^2 + 2.8^2 + 2.9^2 + 4.3^2} = 6.9$$

Figure 1. Mention in the legend the existence of a small water band at about 2050.4 cm⁻¹.

We thank the reviewer for this suggestion, we have added it in the figure caption:

A small water band at 2050.5 cm⁻¹ can interfere with COS at 2050.4 cm⁻¹ and can affect the COS correction for water vapor without a split fit at 2050.45 cm⁻¹ (Sect. 2.3).

Figure 2. Explain why response curves for CO₂ and CO were corrected for drift using a reference cylinder. How important was the drift and how many response curves were corrected. The residuals were averaged isn't it?

The concentrations drifted on the order of 8 ppt for COS, 3 ppm for CO₂ and 5 ppb for CO in these calibration experiments. The response curves shown in Fig. 2 are based on single experiments and the consistency of the response curves when the experiment was repeated is discussed in section 2.2.1. The residuals are indeed shown as averages. In the revised version we added error bars to show the standard deviation of the measurements. We added the following in the legend of Fig 2 to indicate the size of the drift that was corrected for with the reference cylinder:

The drift was on the order of 8 ppt for COS, 3 ppm for CO₂ and 5 ppb for CO and would have given biased results if this was not corrected for.

Figure 7. It is possible to simulate a nocturnal gradient of COS and CO₂ from the assigned cylinder values corresponding to the green and orange samples. During the day the atmosphere is rather well mixed so COS is high and CO₂ is low (green dots). During the night the atmosphere is stratified. Assuming that the ecosystem removes COS from the atmosphere and respire CO₂, COS will decrease (- 41.2 ppt) and CO₂ will increase (+ 9.2 ppm, orange dots). It is difficult to visualize the COS gradient because the signal is very noisy. I suggest to zoom in and check the difference between a series of consecutive orange and green dots. Is the difference about 40 ppt? You can also calculate daily averages and check again that the difference approaches 40 ppt. For CO₂ it looks fine.

We do not understand what the reviewer meant to say here. No ambient concentrations but the cylinder measurements are shown in Fig. 7, which do not represent atmospheric mixing or ecosystem fluxes of COS and CO₂. If the reviewer meant to say that the difference in concentrations between the different cylinders should be constant, that is true. In fact, that is presented in Fig. 9 where one cylinder is used to correct the other cylinders. If any drift would occur in one of the cylinders then it would be visible in these plots.

Figure 8. I think you should compare what is comparable. The leaky valve was removed January 7, 2015 so my suggestion is to redraw the transparent dark histograms using data collected after this date.

The leaky valve has affected the absolute value of the measurements, as we saw in Fig. 7, however, there is no potential influence on the standard deviation of the measurements. We therefore did not change the histograms in Fig. 8.

Figure 9. It seems that the grey dots (upper left panel) have turned to black in the upper right panel.

Actually the right panel only shows data from one cylinder, so there is only one color. In the left panel we calibrate the COS data with a single bias correction. This correction requires one reference cylinder and two of the three cylinders are corrected with this, which are shown as grey and black. In the right panel we calibrate the COS data based on response curves that were determined hourly from two of the three cylinders, and the one remaining cylinder that is corrected with these response curves is shown in black. This is mentioned in the caption of Figure 9 (“For the fixed response curve measurements of two target cylinders are shown, for the hourly response curve only one is shown, as two of the three cylinders were needed to determine the response curve.”). In the revised version we repeat this in the main text.

Figure 10. This is a poorly informative figure. I suggest to focus on diurnal variations and land-ocean differences rather than on seasonal variations.

Figure 11. This is also a poorly informative figure because the new dataset is incomplete (see also general comments above).

See also our general reply and our reply to the first general comment, there we explained that the aim of this manuscript is to study the performance of the instrument, and not to interpret the COS signals. Even though the dataset does not cover a full year, Figure 11 (13 in the revised version) showed us that the QCLS measurements capture the seasonal variation well and that the measurements are comparable with those at other coastal sites at similar latitudes. With this figure we demonstrated the potential of the QCLS analyzer in the field and we can say that the accuracy and precision of the analyzer is suitable for long-term atmospheric monitoring. For these reasons we stick to only show Fig. 11 and 12 (12 and 13 in the revised version). We did add a discussion in section 3.1 about the suitability of the QCLS for different potential purposes, including that of interpreting profile measurements.

Figure 12. The upper left panel should be removed and the upper right panel should be presented after figure 4. I think it is more interesting to display the temporal difference between CRDS and QCLS data, the standard deviation of the difference and to look for temporal trends than to plot one against the other and calculate a slope. Ordinates are missing.

See also our general reply. We decided to move the flask comparison between the QCLS and the GC-MS to the measurement comparison section (3.2 in the revised manuscript) and separated this section in three paragraphs, as presented in a previous reply. The three different measurement comparisons are now introduced so that it is clear that the flask vs in situ comparison at the QCLS is not an instrument comparison.

We agree that it is worth showing the temporal difference between the CRDS and QCLS, also between the QCLS in situ and flask measurements. We replaced the slope curves for the

difference plots, which also include the frequency distribution of the differences plotted as a histogram.

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.

Yes, the mean bias must be compared with the quadrature sum of “Transfer to calibration standards” and “Calibration using calibration standards” (because the target measurements are treated as real ambient air samples), which indeed makes 4.0 ppt for COS, 0.17 ppm for CO₂ and 2.4 ppb for CO. We thank the reviewer for the suggestion to add the standard error, we took this into account in the comparison. We added the following in the discussion of Fig. 8 in Section 3.1.

After the corrections are applied, the mean offset of the measurements is within 3.3 ± 0.2 ppt for COS, 0.05 ± 0.003 ppm for CO₂ and 1.7 ± 0.1 ppb for CO over the period of 35 days. The standard errors indicate that the mean offset is significantly different from 0. Still, the offsets are within the expected uncertainty based on the relevant uncertainties listed in Table 5, namely the transfer to calibration standards and transfer to field measurements of which the quadrature sum is 4.0 ppt COS, 0.17 ppm CO₂ and 2.4 ppb CO. The fact that the mean offsets are within the expected uncertainties indicates that these uncertainties listed in Table 5 properly envelope the uncertainties of the field measurements.

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.

See also our general reply.

The uncertainties presented in Table 5 are the one-second uncertainties, and we would like to keep this consistent. We do agree that the precision on time scales longer than 1 second is also relevant as it reflects the influence of instrument drift. The repeatability also depends on the calibration strategies, e.g. very frequent calibrations would be able to remove drifts on the time scale of the calibration period; however, the use of the calibration gas is not affordable for the field experiments. We would like to stick to one-second uncertainties in Table 5 for consistency, and changed the name “measurement repeatability” into “measurement precision”. Still, we agree that it is relevant to discuss precisions on longer timescales; therefore we added a discussion about this in section 3.1:

Furthermore, in Table 5 and Fig. 7 we present one-second precisions, but also precisions on longer timescales are relevant to reflect the influence of instrumental

drift. Typically, precision depends on the length of the averaging period, where it can either decrease for Gaussian noise or increase for instrument drift. We will see in Fig. 8 (discussed later in this section) that for two-minute averaged target measurements the standard deviation varies between 4.9 and 6.6 ppt. This reflects the typically higher values for precision on longer timescales than one second. We provide the numbers for COS here, but the same holds for CO₂ and CO.

Note that the mean offset in the upper right plot of Fig. 8 is changed from -3.4 to -1.7 in the new version of the manuscript. A wrong cylinder value was used in the calculations for this particular plot.

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 interpret 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 than 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.

The reviewer made a very good point here, and we are thankful for that. We agree that we should discuss the different purposes of the QCLS measurements and the different uncertainties associated with that. Indeed, the measurements do not need to be transferred to the NOAA/WMO scales if only the gradients were to be analysed based on the measurements from different heights. In fact, another reason to transfer our measurements to the NOAA/ESRL scale is that COS in cylinders tends to drift over time, and it is therefore important to achieve good accuracy by linking the measurements to a well-maintained scale. We made a number of changes (additions) in the manuscript to cover this. In the introduction section we introduce the different requirements for different types of analyses:

Besides the difference in requirements for precision between different experimental setups, the type of analyses intended for a dataset also determines the requirements for precision and accuracy of the measurements. If the intention is to compare atmospheric concentrations across sites, then accuracy is important because data

from different sites must be on consistent scales. On the other hand, short-term precision is more important than accuracy when differences between heights are to be interpreted (e.g., as in estimation of fluxes from profile measurements). Following the K -parameterization formulation of the flux-gradient method (e.g. Meredith et al., 2014), $F = -K \frac{\Delta C}{\Delta z} \rho$, the precision required to capture the concentration differences between heights (ΔC) mostly depends on the size of the fluxes F , the height difference Δz , the turbulence conditions, which is represented by the eddy diffusivity K , and to a lesser extent by the molar density of air ρ . To be able to capture COS fluxes of e.g. $10 \text{ pmol m}^{-2} \text{ s}^{-1}$ over a height difference of 20 meters the measurement precision of COS should be better than 0.5 ppt under high turbulent conditions ($K = 10 \text{ m}^2 \text{ s}^{-1}$) and 4.8 ppt under low turbulent conditions ($K = 1 \text{ m}^2 \text{ s}^{-1}$). [...].Furthermore, we evaluate the total uncertainty of the measurements by combining the uncertainties of scale transfer, water vapor corrections and the measurement precision. Based on the precision and accuracy that we derive from these experiments, we discuss the suitability of COS measurements on this instrument for different purposes; that is, for interpreting profile measurements and comparing concentrations across sites.

The following discussion is added in section 3.1 (precision and accuracy):

We have now presented the total uncertainty for long-term concentration monitoring at an atmospheric station, but not all uncertainties are relevant for every type of analyses. If the data are to be compared across different sites, then the data should be on the same scale and the accuracy of the measurements is important. However, if data from the same site and the same instrumentation are compared for example to do flux-gradient analysis, then there is a less stringent need for accuracy and the short-term precision is more important. If the uncertainties related to the transfer of the scales are not taken into account then the total uncertainty would be 5.2 ppt for COS, 0.11 ppm for CO_2 and 1.4 ppb for CO. This includes both the uncertainty of the water vapor correction and the one-second measurement precision. The uncertainty of the water vapor correction cannot simply be ignored as gradients in H_2O can exist as well. We observed that COS concentrations between 40 and 7 m differ on average by 1.7 ± 5.3 ppt during daytime and 11.4 ± 33 ppt during nighttime. Given that the gradients should be larger than the uncertainties mentioned above, the daytime gradients are too small to be able to use the flux-gradient method. The suitability of the flux-gradient method is further dependent on the choice of the measurement height (gradients are larger closer to the surface), the size of the fluxes at a given site and the turbulence conditions. Even though the use of the flux-gradient method is limited by the measurement uncertainty that we found here, COS profile measurements from this instrument can still be useful to derive storage fluxes, as nighttime gradients are typically larger.

Additionally, the water vapor correction uncertainties may also cancel out unless there is a very strong vertical H_2O gradient. While this might be challenging to determine, given the potential for sampling line artifacts, some consideration of this would be worthwhile.

We agree that when H₂O would be similar with height the water correction uncertainty would cancel out, but unfortunately we cannot simply make this assumption. We added a sentence in section 3.1 to mention this:

*If the uncertainties related to the transfer of the scales are not taken into account then the total uncertainty would be 5.2 ppt for COS, 0.11 ppm for CO₂ and 1.4 ppb for CO. This includes both the uncertainty of the water vapor correction and the one-second measurement precision. **The uncertainty of the water vapor correction cannot simply be ignored as gradients in H₂O can exist as well.***

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 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.

See also our general reply and our reply to a comment earlier about short-term precision and repeatability. We agree that also longer-term precisions are relevant in this study, because, indeed, we also discuss hourly and daily averaged measurements in Fig. 10 and 11. As we said in an earlier reply we discussed precisions on timescales longer than one second in section 3.1.

We also agree with the point that the differences between 40 and 60 m can indeed be significant and that this is reflected in the 9.7 ppt standard deviation. Based on this comment we checked if these moments with significant differences affect the flask – in situ comparison and removed one flask – in situ pair from the comparison. For the moments that the flask and in situ measurements are compared the standard deviations are smaller. We added some further discussion about the 40-60 differences:

Although the mean difference of 0.7 ppt is not significant, the 9.7 ppt standard deviation points to actual differences between the 40 and 60 m concentrations. For the moments where the flask and in situ measurements are compared the difference is -4.2 ± 3.4 ppt. We neglected one flask – in situ pair for which the 40 and 60 m difference was larger than 10 ppt. For the remaining data we do not expect a large bias associated with including ongoing results from the 40 m height in the averages.

Specific comments:

Page 2, lines 31-33: Regarding the need to constrain COS to 1.5 to 4.0 ppt to constrain CO₂ fluxes matching a 1 ppm change. It is my understanding that LRU (leaf-scale relative uptake) scales with the ratio of CO₂/COS mole fractions. In this case if CO₂ is 400ppm and COS is 500ppt, then a 1 ppm draw-down in CO₂ 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.

This is a very good point and we thank the reviewer for correcting this detail! Reviewer 1 also asked for a clarification of the calculation from LRU to a COS measurement precision of 0.3-0.8 %. Taking this together we have corrected the sentence such that it now reads as:

*If we were to infer the gross fluxes from chamber measurements with CO₂ measurements better than 1 ppm, then, given the leaf-scale relative uptake ratio of COS/CO₂ 1.5 – 4.0 (Stimler et al., 2010b; Seibt et al., 2010; Berkelhammer et al., 2014), our goal would be to have measurement precisions of COS better than **1.9 – 5.0 ppt for COS (calculated from LRU and scaling with the ratio of COS/CO₂ mole fractions gives e.g. 1.5*(500/400) = 1.9 at the ambient level of 500 ppt COS and 400 ppm CO₂).***

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 CO₂)? 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 longterm vision of this measurement set-up and how COS (and therefore CO₂) fluxes will ultimately be estimated.

A potential type of analysis for this dataset is indeed to use the flux-gradient method. In the revised version we stated more explicitly that the goal is to evaluate and improve the performance of the QCLS and to explore the suitability of the instrument for the different potential analyses. In line with this we also mention the flux-gradient method in the introduction and we discuss the necessary precision of COS to be useful for gradient measurements:

On the other hand, short-term precision is more important than accuracy when differences between heights are to be interpreted (e.g., as in estimation of fluxes from profile measurements). Following the K-parameterization formulation of the flux-gradient method (e.g. Meredith et al., 2014), $F = -K \frac{\Delta C}{\Delta z} \rho$, the precision required to capture the concentration differences between heights (ΔC) mostly depends on the size of the fluxes F , the height difference Δz , the turbulence conditions, which is represented by the eddy diffusivity K , and to a lesser extent by the molar density of air ρ . To be able to capture COS fluxes of e.g. $10 \text{ pmol m}^{-2} \text{ s}^{-1}$ over a height difference of 20 meters the measurement precision of COS should be better than 0.5 ppt under high turbulent conditions ($K = 10 \text{ m}^2 \text{ s}^{-1}$) and 4.8 ppt under low turbulent conditions ($K = 1 \text{ m}^2 \text{ s}^{-1}$).

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?

This is a good point and this was also raised by Reviewer 1. We believe that an extra table is not necessary, but in the revised version we do track the cylinders throughout the manuscript. We added the following sentence in section 2.2.2.

To trace the three working standards that we used in the field we refer to these with numbers #1, #2 and #3 throughout the manuscript. Other working standards that were used in this manuscript but not in the field we refer to as “A”, “B”, “C” and “D”.

We made this consistent throughout the manuscript: in Table 1, 2 and Fig. 5, 6, 8, 9 (numbers referring to the revised manuscript).

This should make clear that the cylinders used for Table 2 are a separate set of cylinders (A, B, C, D) than those in the field (#1, #2, #3). The cylinder numbers are now also given in Fig 8 (Fig 9 in original version of manuscript). Also when we discuss the cylinder drift in section 3.1 we state specifically which cylinder numbers it is associated with.

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

Yes, that is indeed what we mean. The sentence is corrected as suggested.

Page 8, lines 16-17 (and Table 5): Could you provide more flexible uncertainties for the water vapor correction for COS, CO₂, 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 % H₂O. 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 H₂O concentrations (up to 2.1%). I would expect the uncertainty to scale with H₂O, 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 somehow in Table 5 and also reflected in the total uncertainty.

If we give an uncertainty on the slope this would indeed mean that that the uncertainty scales with H₂O and the absolute concentrations. However, it does not appear from Fig. 4 that the uncertainty scales with H₂O. Therefore, we are not in favour of giving an uncertainty on the slope, but rather on the concentration itself, where the uncertainty is determined based on the standard deviation of the residuals of the fit. In the revised version we do discuss in Section 3.1 that the uncertainty scale with the COS, CO₂ and CO concentrations.

The overall uncertainty is calculated as the quadrature sum (square root of the sum of squares) of the individual uncertainties. The overall uncertainty that is presented in Table 5 characterizes the typical measurement uncertainty but may vary depending on the conditions. For example, the measurement precision varies, as we discussed in the previous paragraph, and the water vapor correction uncertainty scales with the COS, CO₂ and CO concentrations.

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?

We agree that this flask analysis in fact shows an instrument comparison, but that we did not discuss it as such, while it adds valuable information. We added the following discussion to cover this:

The flask pairs have a mean deviation from the assigned cylinder values of +3.0 ppt for the QCLS (1.7 ppt when excluding the drifting flask pair) and -3.0 ppt for the GC-MS measurements. Although these deviations are within the measurement uncertainty, the GC-MS measurements are consistently lower than those on the QCLS, with an average difference of 5.1 ppt (excluding the drifting flask pair). We were not able to find an explanation for this bias.

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).

The drift was observed for two of the three cylinders used in Lutjewad, but after the measurement period. Similar drift (2-3 ppt per month) was observed at NOAA/ESRL as well. During the measurement period in Lutjewad we did not see a significant change in concentrations (determined from calibrations) between July 2014 and March 2015. However, we did see drift in two of the three cylinders when they were re-calibrated in November 2015. For the third cylinder we did not observe drift of the same size. We added a more detailed discussion of this in section 3.1:

Additional to the uncertainties presented in Table 5 we have observed that COS can decrease over time in uncoated aluminum cylinders. First, we did not find indications that our calibration standards drifted during field measurements at the Lutjewad station; calibrations in July 2014 and March 2015 showed a decrease of only 2.2 and 1.1 ppt for cylinder #1 and #2 over this 8-month period, which is well within the measurement uncertainty. However, a re-calibration in November 2015 showed a decrease of 18.2 and 24.1 ppt in these cylinders, a decrease with a rate of 2.3 and 3 ppt per month, while cylinder #3 only changed by 1.9 ppt. Cylinder #3 is not different from #1 and #2 (they are all uncoated aluminum cylinders), but cylinder #1 and #2 were stored with a pressure of 25 and 40 bar, which is lower than 130 bar for cylinder #3. We cannot confirm if the drift in the cylinders is related with the cylinder pressure; however, it is the only difference that we were able to find.

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).

We agree that section 3.2 (in the first version) fits better after the measurement comparisons of section 3.3 (in the first version). In the revised version we have moved the measurement comparison section in front of the section that shows the measurement record.

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.

We agree that the reason for doing flask measurements should be better introduced. In fact, multiple reasons that you give are true. For this paper the main reason to do flask measurements is to check the water vapor correction and the comparison with the GC-MS, which indeed needs to be better presented as you also indicated in your comment on Page 11-12, Sec. 2.6. However, testing the QCLS for its suitability to do flask measurements is interesting for the COS community in general, as flask measurements can be useful both during short field campaigns or for longer-term monitoring where in situ measurements from the QCLS are not possible/available. At this moment we monitor COS from the Lutjewad site with flask measurements and use the QCLS for in situ measurements during field campaigns elsewhere. For the revised version we have reconsidered the placement of the flask measurements in the manuscript and have put it into section 3.3 (in the first version, 3.2 in the revised version). We also introduce the flask measurements already in the Introduction section:

Besides in situ measurements, flask or canister measurements can be a valuable tool for providing information about ambient concentrations of COS as well. For example, flask measurements were used before when constructing an historical record from firm air (Montzka et al., 2004), during field campaigns (White et al., 2010; Blonquist et al., 2011) and for long-term monitoring (Montzka et al., 2007). [...] In addition to the experimental setup for continuous in situ measurements we developed a setup to analyze flasks, which we used to make a comparison with GC-MS measurements of flasks and to assess the laboratory derived correction for water vapor interference.

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

No, this value is really -0.02 ppm. This number only compares the QCLS and CRDS measurements from January 7 until January 29, because for that period the TDLwintel correction was turned on and in that period the data needed an extra water correction on top of it (because the TDLwintel correction only is not sufficient). The -0.12 ppm value covers the whole period from January 7 until the end of April. The comparison that we make in lines 4-7 of page 16 is between having the linear water correction curve on top of the TDLwintel water correction and not having this extra water correction curve. With these numbers we want to indicate that the extra correction curve is needed to properly correct for the water effect.

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.

Good point, this is indeed not mentioned. We made a footnote of this in Table 5:

**Average uncertainty over 4 cylinders in Table 2 (method 3).*

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

See also our reply to Page 8, lines 16-17.

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.

Renaming this may indeed help clarifying the meaning of it. In fact, this uncertainty is the calibration of field measurements as done with the calibration standards. In the revised version we renamed “Calibration using calibration standards” to “Measurement calibration”.

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 regression 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.

We like this suggestion and followed on this with something similar; in the revised version we derive the linear regression from averages over every 0.1 % H₂O range. By doing this we made sure that the regression is not influenced by the fact that there are more data points towards lower H₂O concentrations. The slopes derived with this regression are almost identical to those in the first version, but the uncertainty is reduced (e.g. from 3.5 to 2.9 ppt for COS). The slightly different slopes are now used consistently throughout the manuscript.

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.

We agree that this helps clarifying some cylinder logistics and have modified the figure such that now 3 calibration cylinders and 1 nitrogen cylinder are shown in Figure 5. We also added cylinder numbers #1, #2, #3 to repeat that those cylinders were the ones brought to the field.

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).

Good suggestion, in the revised manuscript we discuss this in section 3.2 (which is where these flask measurements are discussed in the new version).

The flask pairs have a mean deviation from the assigned cylinder values of +3.0 ppt for the QCLS (1.7 ppt when excluding the drifting flask pair) and -3.0 ppt for the GC-MS measurements. Although these deviations are within the measurement uncertainty, the GC-MS measurements are consistently lower than those on the QCLS, with an average difference of 5.1 ppt (excluding the drifting flask pair). We were not able to find an explanation for this bias.

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.

This point was also raised by Reviewer 1 and we have made the COS comparison a separate figure. Also we added the instrument comparison between QCLS and GC-MS with flasks in section 3.2 (in the revised version).

M. von Hobe (Editor)

Please also consider the following issues when revising the manuscript:

In Figure 7, the stability improvement after the temperature stabilization measures on March 25 is clearly visible. But while for CO and OCS, the data scatter nicely around the respective assigned cylinder values, the measured CO₂ values are significantly (by about 13 ppm) below the assigned cylinder values. Please explain this in the revised manuscript!

This is indeed the case, and it is because the data in Figure 7 are raw (but hourly averaged) data and are not yet corrected with the instrument response. After this correction the data are close to the assigned cylinder values, e.g. for CO₂ the values are around 405 (orange), 398 (green) and 384 ppm (blue), which becomes 420, 412 and 398 ppm after correcting with the response curve (for CO₂: $1.038x-0.839$, see fig 2). These corrected values are close to the assigned cylinder values. In the revised manuscript we made a note of this in the figure caption:

The data shown here are uncorrected data and are not calibrated with a response curve. The concentrations are therefore not necessarily close to the assigned cylinder values.

In the introduction, precision is defined "as the standard deviation over a two minute period". It would be more useful to look at precision achieved with different averaging times, because i) you have various sources of random noise with different time scales (e.g. temperature fluctuations), and ii) the time scales of the uncertainties are important in the context of different scientific applications of the data (this issue is raised by referee #2). It would be nice to construct and show an Allen plot to show how the random variability varies with time. With the amount of measurements you have made, this should be fairly easy to do.

We agree that also longer-term precisions are relevant in this study, because, indeed, we also discuss hourly and daily averaged measurements in Fig. 10 and 11. We agree that an Allan variance plot can nicely show the capability of averaging the data to reduce noise level; however, we find it less useful in this manuscript where we deal with changing instrument precision over time, which is not represented in Allen plots. In the revised manuscript we did add a discussion in section 3.1 on the uncertainty on timescales of 2 minutes, as obtained from Fig 8 (in the new version).

Furthermore, in Table 5 and Fig. 7 we present one-second precisions, but also precisions on longer timescales are relevant to reflect the influence of instrumental drift. Typically, precision depends on the length of the averaging period, where it can either decrease for Gaussian noise or increase for instrument drift. We will see in Fig. 8 (discussed later in this section) that for two-minute averaged target measurements the standard deviation varies between 4.9 and 6.6 ppt. This reflects the typically higher values for precision on longer timescales than one second. We provide the numbers for COS here, but the same holds for CO₂ and CO.

In the same context, a one-second-precision and an overall uncertainty for one second data of 4.3 ppt and 7.1 ppt respectively are given for OCS. Looking at Figure 7, the scatter for 1-hour-averages looks larger, and it also seems peculiar that the standard deviation for 1 minute averages given in Table 3 should be larger than the standard deviation for one-second-data. Please look at this carefully and explain how you calculated the averages and corresponding standard deviations on different time scales!

The scatter in Fig. 7 is indeed larger than the reported precision and total uncertainty in Table 5. The data from Fig. 7 can be compared with the precision in Table 5 after corrections for instrument drift are applied, which is what we see and evaluate in Fig. 9.

The precision on timescales longer than one second is typically larger due to instrument drift. That explains why the one-minute average standard deviations in Table 3 are larger than the one-second standard deviations in Table 5. In the revised version we discuss the precision on longer time scales in more detail in section 3.1, see also our reply to the previous comment.

Did you ever try to calibrate with higher concentration OCS standards? As far as I know, the OCS standards that can be purchased from NOAA are typically secondary standards (ambient air samples that NOAA measures and certifies). However, in your acknowledgements, you mention the preparation of gravimetric standards. Can you indicate where in your study such gravimetric standards were used? Using one or two standards with higher concentration would certainly make the response curve for OCS shown in Figure 2 even more useful than the two-point-curve.

We had only two NOAA calibrated standards for COS in our lab, as these were the two that we purchased from NOAA. The gravimetric standards are only used at NOAA to calibrate these standards. It would indeed have been nice to calibrate the analyzer with higher (and lower) COS standards, but unfortunately we didn't have the tools available.

For the H₂O correction, where do you obtain quantitative information of the water broadening coefficients? If this information is not published, the experimental derivation should be described.

Unfortunately this work is not published. The information given in lines 1-5 on page 9 regarding the water broadening coefficients estimates from the manufacturer does not add critical information. For these reasons we decided to remove this information from the manuscript.

And I find the discussion somewhat confusing, because on page 9, line 13 you state that "optimized broadening coefficients are equal to 1.0, 2.15 and 1.0 for COS, CO₂ and CO". But a water vapour broadening coefficient of 1.0 would mean that the water vapour effect is the same as for air, so water vapour broadening does not play a significant role, correct? It would be nice to get a quantitative feeling how much the water vapour broadening actually affects the measured mixing ratios...

It is correct that a water broadening coefficient of 1.0 actually has no effect. The uncertainties of the water broadening coefficients actually tell more about the significance of these numbers. For the revised manuscript we discuss the uncertainties of the water broadening coefficients in Section 2.3:

As the optimization of the broadening coefficients depends on the curves in Fig. 4 (most optimal is when the slopes are equal to zero), the uncertainty of the water correction also entails an uncertainty in the broadening coefficients. We find that the uncertainties of the broadening coefficients are equal to 0.5 (COS), 0.03 (CO₂) and 0.7 (CO). This means that varying the broadening coefficient of COS from 1.0 to 1.5 only changes the COS concentration by 2.9 ppt (at a concentration of 450 ppt COS). The large uncertainties of the COS and CO broadening coefficients therefore indicate that these coefficients have relatively little effect on the concentrations.

I understand the numbers given as $\frac{\Delta C}{C}$ in Section 2.5 in a way that you reduced the response of your system to ambient temperature variability? Was it not possible to stabilize to a fixed temperature? This seems not so difficult...

Yes, it is correct that we tried to reduce the response to ambient temperature variability and indeed it would have been most optimal to stabilize the temperature, either the lab room where the system is placed, or the temperature of the system itself. Temperature-stable room conditions were very difficult to achieve, especially for field conditions where it is difficult to arrange temperature-controlled conditions. We did efforts to control the temperature in the room where the instrument was placed but only ended up with quick temperature changes from an air conditioning. The mean temperature was then constant at 25 °C, but with fluctuations of ± 0.5 degrees, which only resulted in larger instrument drift. We then tried to stabilize the instrument itself by extending the cooling loop and adding isolation. We did not manage to further control the temperature of the system itself with the given time and availability of the instrument before later field work started. However, there are new ideas in the community to further stabilize the instrument. We encourage these developments and may follow these in the future.

Continuous and high precision atmospheric concentration measurements of COS, CO₂, CO and H₂O using a quantum cascade laser spectrometer (QCLS)

Linda M.J. Kooijmans¹, Nelly A.M. Uitslag¹, Mark S. Zahniser², David D. Nelson², Stephen. A. Montzka³, Huilin Chen^{1,4}

¹Centre for Isotope Research (CIO), University of Groningen, Groningen, The Netherlands

²Aerodyne Research Inc., MA, USA

³NOAA Earth System Research Laboratory, Boulder, Colorado, USA

⁴Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA

10

Correspondence to: Huilin Chen (Huilin.Chen@rug.nl)

Abstract. Carbonyl sulfide (COS) has been suggested as a useful tracer for Gross Primary Production as it is taken up by plants in a similar way as CO₂. To explore and verify the application of this novel tracer, it is highly desired to develop the ability to perform continuous and high precision in situ atmospheric measurements of COS and CO₂. In this study we have tested a quantum cascade laser spectrometer (QCLS) for its suitability to obtain accurate and high precision measurements of COS and CO₂. The instrument is capable of simultaneously measuring COS, CO₂, CO, and H₂O after including a weak CO absorption line in the extended wavelength range. An optimal background and calibration strategy was developed based on laboratory tests to ensure accurate field measurements. We have derived water vapor correction factors based on a set of laboratory experiments, and found that for COS the interference associated with a water absorption line can dominate over the effect of dilution. This interference can be solved mathematically by fitting the COS spectral line separately from the H₂O spectral line. Furthermore, we improved the temperature stability of the QCLS by isolating it in an enclosed box and actively cooling its electronics with the same thermoelectric chiller used to cool the laser. The QCLS was deployed at the Lutjewad atmospheric monitoring station (60 m, 6°21'E, 53°24'N, 1 m a.s.l.) in the Netherlands from July 2014 to April 2015. The QCLS measurements of independent working standards while deployed in the field showed a mean difference with the assigned cylinder value within 3.3 ppt for COS, 0.05 ppm for CO₂ and 1.7 ppb for CO over a period of 35 days. The different contributions to uncertainty in measurements of COS, CO₂ and CO were summarized and the overall uncertainty was determined to be 6.9 ppt for COS, 0.21 ppm for CO₂ and 3.4 ppb for CO for one second data. A comparison of in situ QCLS measurements with those from concurrently filled flasks that were subsequently measured by the QCLS showed a difference of -9.7 ± 4.6 ppt for COS. Comparison of the QCLS with a cavity ring-down spectrometer showed a difference of 0.12 ± 0.77 ppm for CO₂ and -0.9 ± 3.8 ppb for CO.

1. Introduction

Carbonyl Sulfide (COS) has been suggested as a potential tracer for photosynthetic CO₂ uptake (Sandoval-Soto et al., 2005; Montzka et al., 2007; Campbell et al., 2008; Berry et al., 2013, Asaf et al., 2013), as it follows the same uptake pathway into plants through stomata as CO₂, but is not generally re-emitted by plants (Protoschill-Krebs and Kesselmeier, 1992; Protoschill-Krebs et al., 1996; Stimler et al. 2010b). COS therefore provides a means to partition Net Ecosystem Exchange into Gross Primary Production (GPP) and Respiration. As large uncertainties in the COS budget remain, field measurements of COS and CO₂ concentrations and fluxes from leaf- to ecosystem- and regional scale are required for the COS tracer method to be tested and validated (Wohlfahrt et al., 2012; Berkelhammer et al., 2014). Therefore, there is a need for high frequency and high precision measurements techniques of COS and CO₂.

10

Several past studies on COS have relied on discrete (flask) samples analyzed with Gas Chromatographic Mass Spectrometry (GC-MS; Montzka et al., 2007; Stimler et al., 2010b; ~~Belviso et al., 2013~~). For example, the global atmospheric flask sampling network described by Montzka et al. (2007) has allowed a foundation for understanding COS concentrations over annual cycles on global scale. ~~Although the GC-MS technique can be used for *in situ* measurements (Miller et al., 2008; Belviso et al., 2013), this technique does not typically allow for high frequency measurements of 1 to 10 Hz. for continuous high frequency measurements and is less appropriate for providing insight in, for example, diurnal variations of local fluxes from soil and plants.~~ Recent developments of Quantum Cascade Laser spectrometers (QCLS) have enabled *in situ* trace gas measurements including COS. These instruments have proven to be a valuable tool for continuous high frequency measurements of COS and CO₂ up to a frequency of 10 Hz (Stimler et al., 2010a; 2010b; Asaf et al., 2013; Commane et al., 2013; Berkelhammer et al., 2014; Maseyk et al., 2014; Commane et al., 2015).

20

The required measurement precision (in this study we define precision as the standard deviation over a two minute period) for studies of exchange processes of COS and CO₂ between biosphere and atmosphere depend on the concentration change that these gases undergo in any given experiment. On the regional scale, COS shows seasonal variations typically between ~100 to 150 ppt at continental sites in the Northern Hemisphere (NH), and 40 to 70 ppt in the Southern Hemisphere (SH) and at marine sites (Montzka et al., 2007). CO₂ seasonal variations typically reach up to 15 ppm on the NH and are as low as 2 ppm at South Pole (Zhao and Zeng, 2014). For the leaf scale, COS and CO₂ concentration changes can be substantially larger; for example, Berkelhammer et al. (2014) showed that during branch bag measurements COS generally decreased by 180 to 240 ppt during active photosynthesis and CO₂ concentrations can easily change by 200 ppm, depending on the setup.

25

30

~~Besides the difference in requirements for precision between different experimental setups, the type of analyses intended for a dataset also determines the requirements for precision and accuracy of the measurements. If the intention is to compare atmospheric concentrations across sites, then accuracy is important because data from different sites must be on consistent scales. On the other hand, short-term precision is more important than accuracy when differences between heights are to be~~

interpreted (e.g., as in estimation of fluxes from profile measurements). Following the K-parameterization formulation of the flux-gradient method (e.g. Meredith et al., 2014), $F = -K \frac{\Delta C}{\Delta z} \rho$, the precision required to capture the concentration differences between heights (ΔC) mostly depends on the size of the fluxes F , the height difference Δz , the turbulence conditions, which is represented by the eddy diffusivity K , and to a lesser extent by the molar density of air ρ . To be able to capture COS fluxes of e.g. $10 \text{ pmol m}^{-2} \text{ s}^{-1}$ over a height difference of 20 meters the measurement precision of COS should be better than 0.5 ppt under high turbulent conditions ($K = 10 \text{ m}^2 \text{ s}^{-1}$) and 4.8 ppt under low turbulent conditions ($K = 1 \text{ m}^2 \text{ s}^{-1}$). If we were to infer the gross fluxes from chamber measurements with CO_2 measurements better than 1 ppm, then, given the leaf-scale relative (LRU) uptake ratio of COS/CO_2 1.5 – 4.0 (Stimler et al., 2010b; Seibt et al., 2010; Berkelhammer et al., 2014), our goal would be to have measurement precisions of COS better than 1.9 – 5.0 ppt for COS (calculated from LRU and scaling with the ratio of COS/CO_2 mole fractions gives e.g. $1.5 \cdot (500/400) = 1.9$ at the ambient level of 500 ppt and 400 ppm CO_2).

Measurement instruments for long-term atmospheric trace gas concentration monitoring need to meet different requirements than, for example, eddy-covariance measurements. The eddy-covariance technique requires high frequency data (>10 Hz), which typically adversely affects the precision of the measurements compared to 1 Hz data, and requires an averaging period of about 10 to 30 minutes. In contrast to the high frequency required for eddy-covariance measurements, lower frequency measurements (1 Hz) provide useful results over extended measurement periods and enhance the precision of any individual measurement. Furthermore, measurements for long-term monitoring do not require fast response, thus it is not necessary to operate the instrument at high flow rates. As a matter of fact, low flow rates are preferred so that working standards can be used over a long period. This reduces the additional logistics needed for calibration gases, such as filling, calibration and transportation of the standards (Xiang et al., 2014). Besides *in situ* measurements, flask or canister measurements can be a valuable tool for providing information about ambient concentrations of COS as well. For example, flask measurements were used before when constructing an historical record from firm air (Montzka et al., 2004), during field campaigns (White et al., 2010; Blonquist et al., 2011) and for long-term monitoring (Montzka et al., 2007). In this research we developed a robust setup for high precision and long-term monitoring of ambient concentrations of COS, CO_2 , CO and H_2O at different heights from the Lutjewad monitoring station in Groningen, The Netherlands. To this end we employed a ‘QCL Mini Monitor’ from Aerodyne Research Inc. (Billerica, MA, USA) that can operate autonomously and requires little operator attention. We designed an optimal strategy for ‘zero’ air spectral correction and calibration for accurate measurements and we assessed the correction for water vapor interference. In this manuscript we aim to evaluate and improve the performance of the instrument. We will show the precision and accuracy of the instrument with over half a year of field-data and measurements of working standards, and we compare the measurements with other instrumentation. Furthermore, we evaluate the total uncertainty of the measurements by combining the uncertainties of scale transfer, water vapor corrections and the measurement precision. Based on the precision and accuracy that we derive from these experiments, we discuss the

suitability of COS measurements on this instrument for different purposes; that is, for interpreting profile measurements and comparing concentrations across sites. In addition to the experimental setup for continuous *in situ* measurements we developed a setup to analyze flasks, which we used to make a comparison with GC-MS measurements of flasks and to assess the laboratory derived correction for water vapor interference.

5 2. Experimental setup

Before the actual deployment of the instrument in the field we performed laboratory tests to assess the accuracy and traceability of the QCLS measurements and to develop procedures for applying corrections as needed. Here we describe the laboratory tests and we give detailed information about the instrumentation and field setup.

2.1 Instrumentation

10 The ‘QCL Mini Monitor’ that we use is a Tunable Infrared Laser Direct Absorption Spectrometer (TILDAS) using a single continuous-wave quantum cascade laser (Alpes Lasers), which is cooled with a Peltier element to -19.8°C , and using a single photodiode infrared detector (Teledyne Judson Technologies; McManus et al., 2010). The waste heat from both the laser and detector is removed with a recirculating mixture of water containing 25% ethanol, which is temperature controlled with a thermoelectric chiller, ThermoCube 300 (Solid State Cooling Inc, USA). The instrument was initially set to simultaneously
15 measure COS, CO_2 and H_2O at wavenumbers 2050.397, 2050.566 and 2050.638 cm^{-1} , respectively. We extended the range of the laser current to include measurements of CO at 2050.854 cm^{-1} . Figure 1 shows the simulated transmission spectrum of ambient concentrations of COS, CO_2 , CO and H_2O as obtained through the HITRAN 2012 database (Rothman et al., 2013). The precision and accuracy of the measurements will be discussed in Sect. 3.1.

20 The instrument consists of a 0.5 L astigmatic Herriot style multipass absorption cell (McManus et al., 2010) with an effective path length of 76 m. The cell has a temperature between 20 and 24°C , depending on the room temperature and the temperature setting of the thermoelectric chiller. The cell is kept at a constant pressure of 53.3 hPa (40 Torr) with an inlet valve that is controlled by the TDLWINTTEL program (Aerodyne Research Inc., Billerica, MA, USA) based on the measured cell pressure. The same software manages the data acquisition and spectral analysis (Nelson et al., 2004) and calculates dry
25 air mole fractions real-time (1 Hz) through nonlinear least square spectral fits combined with the measured cell temperature and pressure, a constant path length, and the HITRAN 2012 database cross-sections as a function of wavelength. The spectral fit for CO is separated from the fit for COS, CO_2 and H_2O as there is slight interaction of the CO peak with a second absorption line of COS. The COS fit close to the CO peak is linked to the COS peak at lower wavenumbers to improve the fitting for CO. This is achieved by fitting the spectra in two steps: first the mole fractions are determined for both COS peaks
30 independently, second the CO concentration is recalculated with the fixed COS concentration derived from the separated COS peak in the first step.

The TDLWINTEL software has the option to store raw spectra. These spectra can later be used for re-analysis using the so called ‘Playback’ mode of the software. The spectral parameters (line shape and position) for the fits are taken from the HITRAN database (Rothman et al., 2013). The sample spectra are normalized with a ‘zero’ air spectrum to remove background spectral structures and to remove absorbance external to the multi-pass cell (Stimler et al., 2010a; Santoni et al., 2012). The ‘zero’ air spectrum is periodically determined when the cell is flushed with high-purity nitrogen (99.99999%), which we will now refer to as ‘background’ measurement. The nitrogen is first passed over a gas purifier (Gatekeeper, CE-500K-I-4R) to remove CO that is often found in such nitrogen cylinders. The frequency of the laser is locked based on the spectrum measurement of the high strength CO₂ line at 2050.566 as shown in Fig. 1. For automatic start-up, a gas sealed in an aluminum reference cell can be flipped into the optical beam. The reference cell was filled with 8 hPa (6 Torr) COS and 27 hPa (20 Torr) CO. Initially, we could use the peak position of COS in the reference cell to determine the frequency of the laser. However, COS did not last longer than a few months in the reference cell so thereafter the laser frequency was locked only based on the peak position of CO, which did not impact the results.

2.2 Calibration strategy

To allow comparison of QCLS measurements with other instrumentation and across different sites requires traceability to a primary scale. Laboratory tests were conducted to characterize the response of the instrument against ambient air standards from NOAA/ESRL, which were subsequently used to transfer the calibration scale to working standards. Moreover, we performed tests to understand the frequency required for background and reference measurements to ensure reliable and accurate results.

2.2.1 Instrument response

To characterize the response of the instrument for COS, CO₂ and CO mole fractions we used six aluminum cylinders that were calibrated at the NOAA Boulder laboratories on NOAA-2004 COS scale, WMO-X2007 CO₂ scale (Zhao and Tans, 2006) and WMO-X2004 CO scale (Novelli et al., 2003). Five of the six standards were calibrated for CO₂, four for CO and two for COS. The internal surfaces of the COS calibrated cylinders were Aculife-treated. Figure 2 shows that the response curves are linear for CO₂ (between 354 and 426 ppm) and CO (between 94 and 384 ppb). Furthermore, the residuals of these curves do not show signs of non-linearity. For COS conclusions about the linearity of the response are not possible given that we had only two calibrated standards. As the response is linear for CO₂ and CO, we assume that the response is linear for COS as well. The NOAA/ESRL calibrated standards cover a wide range of CO₂ and CO concentrations and response curves that were determined for these gases were consistent when the experiment was repeated. The factor 1.038 to transfer the QCLS CO₂ measurements to the WMO scale is slightly different from the factor 1.05 found by Commane et al., 2013. For COS the concentration range of the NOAA/ESRL calibration standards was smaller, between 447.8 and 486.6 ppt, and combined with the lower precision of these measurements, the response curve has large uncertainty. Therefore, repeating the

experiment resulted in varying response curves, with an average slope equal to 0.99 and the minimum and maximum slope equal to 0.87 and 1.04 respectively. The response curve shown in Fig. 2 is the curve where the two NOAA/ESRL calibration standards measurements showed the lowest standard deviations: 2.7 and 3.0 ppt, with the average standard deviation over different experiments equal to 4.4 ppt. In the next section we will discuss different calibration methods to deal with the higher uncertainty of the COS response curve. Furthermore, we tested the stability of the response of COS, CO₂ and CO with hourly measurements of three working standards over a period of 35 days during field measurements. The results of this will be shown in Sect. 3.1.

2.2.2 Working standards

In this study we use working standards to represent high-pressure cylinders that are calibrated against NOAA or WMO standards in our laboratory. The working standards are used for two purposes: (1) to correct for instrument drift during field measurements and link the measurements to the NOAA or WMO scales, in this case we refer to these standards as reference standards, and (2) to assess the accuracy of the measurements, in this case we refer to these standards as target standards. The working standards used in this study are aluminum gas cylinders (Luxfer, max. 200 bar) filled to high pressure with ambient air at the Center for Isotope Research of the University of Groningen using an oil free air compressor (RIX SA-3) and are used in combination with two-stage regulators (Scott Specialty Gases, model 14). The working standards differ from the NOAA/ESRL calibration standards in the sense that they are uncoated. To trace the working standards that we used in the field we refer to these with numbers #1, #2 and #3 throughout the manuscript. Other working standards that were used in this manuscript but not in the field we refer to as “A”, “B”, “C” and “D”. Using the linear regression curve that we found in Fig. 2, we determined mole fractions in these working standards by considering response curves derived from measurements of the NOAA/ESRL calibration standards. Results of calibrations of three working standards with the QCLS are shown in Table 1 for CO₂ and CO. We calibrated the same working standards with a cavity ring-down spectrometer (CRDS; Picarro Inc. model G2401-m) using the same standards linked to the WMO scale. Calibrations with the QCLS for CO₂ and CO agree with calibrations from the CRDS within their uncertainties, which gives confidence in our calibration method. For COS we could not compare our calibrations with that from other instrumentation.

As we saw in Sect. 2.2.1, the response curve of COS is difficult to determine due to the lower instrument precision and the narrow COS concentration range of the NOAA/ESRL calibration standards. These factors introduce uncertainty in assigning values to the working standards that will be used on-site, especially for those having COS mole fractions outside of this range. Besides that, the number of available calibrated standards used to transfer the scales to other cylinders may be limited in labs, especially for COS, as this gas is usually not one of the standard measured species. The question is therefore what a suitable method is to transfer the scale to the working standards for COS. To test this, we re-analyzed calibration measurements in different ways: (1) with response from the two NOAA/ESRL calibration standards and the curve forced

through a zero-point, (2) with two NOAA/ESRL calibration standards and the curve *not* forced through zero, and (3) with a single bias correction using a NOAA/ESRL calibration standard that has the concentration closest to the working standard.

Table 2 shows the assigned values on the working standards considering the different approaches. The calibration measurement was repeated three times and results are here shown as the average over the three measurement. The fact that the cylinder values need to be extrapolated causes larger deviations compared to when the curve is forced through a zero point (see e.g. cylinder “A” in Table 2). Therefore, when working standards have concentrations outside the range covered by the NOAA/ESRL calibration standards, method 1 is preferred because it avoids extrapolation of the calibration curve by using the zero-point. However, as we saw in Sect. 2.2.1 it is difficult to accurately determine a calibration curve due to the lower precision of the measurements and the narrow COS concentration range of the NOAA/ESRL calibration standards, which holds for both method 1 and 2. In this study we therefore calibrate cylinders and field measurements under the assumption that the data follow a calibration curve with a slope equal to 1 (in Sect. 2.2.1 we saw that the average slope of different calibration experiments was equal to 0.99) and we apply a single bias correction, which is applied in Table 2 as method 3. Theoretically, methods 1 and 2 would give the results closest to reality, here we observe that the difference of method 3 against methods 1 and 2 is small (on average 1.3 ppt) and the result is often in between those of methods 1 and 2. We therefore consider the single bias correction as a good compromise when a calibration curve cannot accurately be determined.

2.2.3. Background and reference strategy

Background measurements with dry nitrogen are required to reduce the effect of curvature of the baseline spectra and are typically done every 2, 5 or 30 minutes by other users of similar QCLS analyzers (Stimler et al., 2010a; Commane et al., 2013). When we tested the required frequency of background measurements we found that COS concentrations can shift to another level, either up or down, uncorrelated with instrument parameters such as temperature, and even when the background measurements were done shortly after each other (every 10 minutes). In a test where we measured cylinder air over a period of five hours, alternated with background measurements every 10 minutes, we found that on average the COS concentration shifts by 5 ppt after every background measurement, but in 32 % of the cases the shift is larger than 10 ppt, with peaks up to 37 ppt. We could not correlate these shifts to changes in instrument temperature or inlet pressure, or the length of the background measurement itself. As we could not find correlations with any parameter, it is unknown what the reason is for these shifting concentrations after background measurements. To be able to correct for the concentration shifts for accurate concentration measurements requires measurements of reference gases at least once within every background cycle. Therefore, increasing the background frequency automatically leads to larger use of a reference cylinder gas. As we also need to consider cylinder logistics (filling, calibrating and transportation) we decided to do less frequent background measurements (once every six hours) with reference cylinder measurements in between. Lowering the background frequency to six hours did not negatively affect the measurement precision, but the accuracy of the measurements is affected due to

instrument drift over the period in between the background measurements. The frequency of reference cylinder measurements that is needed to correct for the instrument drift depends on the rate with which the instrument temperature changes. Hourly measurements of reference standards in the field (during a later field campaign in Hyytiälä, Finland, not shown here) showed that concentrations drift on the order of 100 ppt (COS), 2 ppm (CO₂) and 10 ppb (CO) per degree change of the electronics temperature in a few hours. To test the frequency of reference measurements that is needed to remove drift due to temperature changes, we measured cylinder air over a period of 19 hours, alternated with a reference gas every 15 minutes. During this measurement period the electronics temperature changed by 0.1 °C due to changes of the room temperature, where the most rapid change was 0.06 °C in 30 minutes. Without correcting the data, the standard deviation of the minute averaged data is 11.8 ppt for COS, with drift up to 40 ppt. Correcting the data with reference measurements from every 15 minutes lowers the standard deviation to 4.6 ppt, correcting only every 30 and 60 minutes gives a standard deviation of 5.3 and 9.6 ppt respectively. See Table 3 for an overview of the standard deviations of COS, CO₂ and CO. Based on these results we can say that reference measurements every 30 minutes are sufficient to remove drift within 5.3 ppt with temperature changes up to at least 0.06 °C per 30 minutes. Field measurements have shown that with corrections from 30 minute reference measurements, drift is still sufficiently removed with temperature changes up to 0.2 °C per 30 minutes. Furthermore, improving the temperature stability of the instrument could reduce the drift such that reference measurements are needed less frequently. The effect of improved temperature stability was tested during field operation of the QCLS at the Lutjewad station (see Sect. 2.5).

2.3 Water vapor interference correction

The concentration measurements of gases determined through light absorption spectrometry can be affected by water vapor in the sample air in two ways: (1) by spectroscopic effects (enhanced pressure broadening or direct spectral interference), which will directly modify the absorption spectrum and (2) through dilution of the sample air, which linearly depends on water vapor concentrations. The water vapor interference can be prevented by drying the air before measurement. However, this requires adding a drier to the sampling inlet lines and depending on the drier it can require additional maintenance, which is not favorable for unattended autonomous measurements. As the QCLS includes measurements of H₂O, water vapor interferences can be accurately accounted for in the calculation of dry air mole fractions of the other gases. We determined the water vapor dependence of mole fractions reported by the instrument (TDLWINTTEL software) for COS, CO₂ and CO from laboratory experiments. To do this we measured dry and humidified cylinder air alternately. We humidified the cylinder air with wet silica gel in a filter, giving a water vapor profile from 2.1 % down to 0.2 % H₂O in the sample air when the line was flushed (Rella et al. 2013). The interaction of silica gel with COS, CO₂ and CO was tested by drying the air with magnesium perchlorate or a cryogenic system after the air had passed the wet silica gel. No interaction with COS, CO₂ and CO was found for silica gel, magnesium perchlorate and the cryogenic system: the average difference between the unaffected and humidified-dried silica gel air was maximum 1.0±1.9 ppt for COS, 0.11±0.13 ppm for CO₂ and 0.3±0.4 ppb for CO. Furthermore, we found that a 0.3 nm molecular sieve, which is commonly used to remove water vapor, removes all

COS from cylinder air. The dry air cylinder measurements were used as a reference during the experiment to account for instrumental drift. Figure 3 shows the mole fractions of the humidified air measurements during the experiment with the TDLWINTTEL water vapor correction turned off. Figure 4 shows how the wet/dry ratios of the COS, CO₂ and CO concentrations relate to H₂O when the TDLWINTTEL water correction was turned off (blue). For this figure we combined 5 three water vapor tests, of which one of the three was done with the TDLWINTTEL water correction turned off, the other two had the TDLWINTTEL water correction turned on. We ran the Playback mode of the TDLWINTTEL software to get the data of these water vapor experiments for the case that the correction was turned off, when the real time data were obtained with the correction on. For CO only two of the three experiments are used in the analysis because the third showed larger scatter (colored in gray), indicating instability of the QCLS. We will further discuss the precision of the measurements and its 10 variation over time in Sect. 3.1. The linear regression is derived from averages over every 0.1 % H₂O range such that the regression is not influenced by the fact that there are more data points towards lower H₂O concentrations. Figure 4 shows that the COS, CO₂ and CO wet/dry ratios are all linearly dependent on H₂O. When wet air mole fractions are measured (that is, the TDLWINTTEL water vapor correction is turned off), these curves could act as a water vapor correction factor to obtain dry air mole fractions. The effect of H₂O on the species is +3.0 % for COS, -1.46 % for CO₂, and -0.8 % for CO per % H₂O. 15 The standard deviation of the residuals of the fit translate to an uncertainty in concentrations of 2.9 ppt for COS (at a concentration of 450 ppt), 0.10 ppm for CO₂ (at 400 ppm), and 1.1 ppb for CO (at 150 ppb). These uncertainties are depending on the COS, CO₂, and CO concentrations. The fact that CO₂ and CO show an inverse correlation with H₂O indicates that these species are primarily affected by the dilution effect of water vapor. To the contrary, COS shows a positive relation with H₂O. The reason for this positive correlation is that the baseline shape of the spectra is distorted by a 20 small H₂O peak on the left of the larger H₂O peak to the right of the large CO₂ peak (see Fig. 1). A potential solution to this would be to split the fit between the COS peak and the small H₂O peak, which we will now refer to as “split fit” (in contrast to the “standard fit”). In Fig. 4 the results with the split fit are shown in green. For CO₂ and CO the green curve does not differ significantly from the blue curve, as would be expected. For COS the correlation with H₂O is now negative, with an effect of -0.8 % per % H₂O. A slope of -1 % would be expected due only to pure dilution; however, pressure broadening by 25 water vapor affects the mole fractions as well. Within TDLWINTTEL the width of the peaks are fixed based on the line shape information from the HITRAN database which is only for air-broadening with N₂ and O₂. Water broadening can be greater than air broadening. To correct for the pressure broadening effect the software modifies the width of the absorption line through the so called air “broadening coefficients” which are specific for every spectral line of a certain species. The later version of TDLWINTTEL can also use the water broadening effect by increasing the air broadening coefficients from 30 HITRAN. The residual error of the fit, which is caused by the broadened absorption line but not properly adjusted line width, affects the mole fraction and thereby the slope of the curves in Fig. 4. This can explain why the slopes differ for the different species. Besides that, the measured H₂O concentration on itself has an uncertainty, which contributes to the deviation of the slope from -1 %. For the spectral lines that we use the ratio of the water broadening to air broadening coefficients were estimated by the manufacturer to be 1.5, 1.7 and 1.5 for COS, CO₂ and CO respectively for application of

the standard fit. The slopes that were determined with these TDLWINTTEL corrections (not shown here) were equal to +3.0 % for COS, -0.19 % for CO₂ and +0.1 % for CO per % H₂O without using the split fit. These results show that, even after correcting for water vapor interference by the TDLWINTTEL software, the mole fractions are still water vapor dependent.

5 Using the Playback mode we tried to find the most optimal water broadening coefficients to sufficiently correct for water vapor using TDLWINTTEL. We did this for the three different water vapor tests and for both the standard fit and the split fit. For the standard fit we could not find optimized broadening coefficients for COS because turning the TDLWINTTEL correction on caused an opposite correction and resulted in larger deviations from the assigned cylinder value due to the effect of the H₂O peak on the baseline. For the *standard* fit, the optimized broadening coefficients varied between 2.1 and
10 2.2 for CO₂ and between 1.0 and 2.0 for CO for the different experiments. For the *split* fit we did find optimized coefficients for COS between 1.0 and 1.4 for the different experiments, for CO₂ and CO the same results were found as for the standard fit. When the different experiments were combined the optimized broadening coefficients are equal to 1.0, 2.15 and 1.0 for COS, CO₂ and CO respectively, where the values for CO₂ and CO can be used for both the standard and split fit, and the value for COS is only suitable for the *split* fit. As the optimization of the broadening coefficients depends on the curves in
15 Fig. 4 (most optimal is when the slopes are equal to zero), the uncertainty of the water correction also entails an uncertainty in the broadening coefficients. We find that the uncertainties of the broadening coefficients are equal to 0.5 (COS), 0.03 (CO₂) and 0.7 (CO). This means that varying the broadening coefficient of COS from 1.0 to 1.5 only changes the COS concentration by 2.9 ppt (at a concentration of 450 ppt COS). The large uncertainties of the COS and CO broadening coefficients therefore indicate that these coefficients have relatively little effect on the concentrations. The uncertainty of
20 these optimized broadening coefficients is influenced by the uncertainty of the curves in Fig. 4 and the variation between the different experiments.

Different water vapor correction strategies can now be considered and are summarized in Table 4. An appropriate direct water correction for COS is not possible with the combination of the standard fit and the TDLWINTTEL correction on.
25 However, a correction curve can still be applied to these data with a curve that is determined with the same broadening coefficients as is used for the original data to be corrected. Here we found that for a broadening coefficient of 1.5 with the standard fit the slope of the curve for COS is equal to 0.031. We continued to test the performance of the water vapor correction based on the standard fit and the TDLWINTTEL correction off. We applied the correction curves to field measurements over a period of 35 days in March and April 2015. In Sect. 3.2 we will compare the dry air mole fractions
30 with measurements of a collocated CRDS for CO₂, CO and H₂O and with dry air flask sample measurements for COS.

2.4 Flow schematics for measurements at the Lutjewad station

In July 2014 we deployed the QCLS in the field for measurements at the Lutjewad monitoring station in the Netherlands (60 m, 6°21'E, 53°24'N, 1 m a.s.l.). We use three diaphragm pumps (KNF N-86) to keep the inlet lines well flushed between the

inlet at the tower and the laboratory where the analyzer is positioned (up to 60 m length in the tower and 30 meter from the tower to the lab-building) with a flow of 2 L min^{-1} . Just in front of the pumps we split the line with a tee junction to get a subsample from the tower lines. These subsamples are pulled through the sample cell with an oil-free dry scroll pump (Varian TriScroll) downstream of the cell of the QCLS. 1/2" Synflex (Decabon) tubing is used in the tower and in front of the KNF pumps. We have tested the interaction of Teflon and Synflex with COS: no significant differences in COS mole fractions were found when the air flow was alternately passed over stainless steel and Teflon (0.9 ± 1.9 ppt) or Synflex (0.7 ± 2.6 ppt). We did observe COS production from the KNF N-86 pumps, however, this did not cause problems for the measurement sampling as these pumps were placed on the bypass lines. We use a $5.0 \mu\text{m}$ Teflon filter at the inlet of the tower sample lines to prevent dust, sand and salt to come into the lines. In front of the analyzer we use another $0.5 \mu\text{m}$ stainless steel filter (Swagelok) to prevent pollution of the sample cell. A check valve is placed between the analyzer and the vacuum pump to prevent unfiltered room air to enter the cell **through the pump** in case the vacuum pump suddenly stops. A needle valve was placed between the cell and the vacuum pump to control the flow of the system at 0.16 L min^{-1} , resulting in a 90% response time of ~ 15 s. We use a multi-position Valco valve (VICI; Valco Instruments Co. Inc.) to switch between sample lines from different heights in the measurement tower and to working standards. The Valco valve is controlled by the TDLWINTTEL software in which one can set the sequence and duration of the valve ports to be measured. An interval is set for automatic hourly repetition of the sequence. Until January 2015 we used a solenoid valve (Parker) to switch between dry nitrogen and cylinder or ambient air. However, we found that this solenoid valve was leaking, and was thereby diluting the cylinder and ambient air measurements with dry nitrogen. On January 7, 2015 we changed the setup such that the Valco valve also controlled switching to the dry nitrogen. The Valco valve, KNF pumps, air purifier and $0.5 \mu\text{m}$ filter are built into a 19" rack with 1/8" stainless steel tubing.

For our setup, every hour starts with a measurement from a reference and target standard (3 minutes each). Subsequently, the system alternates between three measurement heights where every height is measured for 8 minutes, meaning that every height is measured two times in an hour. The reference standard was measured every half hour to remove instrument drift. In March 2015 we measured an extra reference gas once every hour, which was used to test the stability of the instrument response over a period of 35 days (the results will be shown in Sect. 3.1). Background measurements were done every six hours with dry nitrogen over 60 seconds. Before the actual background measurement is done, the cell is first flushed for 2.5 minutes to make sure that water vapor is removed from the cell by 99%.

2.5 Temperature stability

We noticed that under lab conditions, when the temperature is controlled within $\sim 0.2 \text{ }^\circ\text{C}$, the precision of the instrument was typically better than with the highly changing temperatures up to $2 \text{ }^\circ\text{C}$ at the measurement station during the course of the day. These temperature dependencies were also observed with other QCLS analyzers in Xiang et al. (2014) and Berkelhammer et al. (2014). Before we made any modifications, the electronics temperature (T_{electr}) varied with $0.91 \text{ }^\circ\text{C}$ with

every degree of changing room temperature (T_{room}), the cell temperature (T_{cell}) varied with $0.11\text{ }^{\circ}\text{C}/^{\circ}\text{C}$. Xiang et al. (2014) showed the potential to improve the temperature stability with an active temperature control using an Oasis 3 chiller (Solid State Cooling Inc, USA) of which the set point can be controlled with the TDLWINTEL software. They improved the T_{cell} variability of $0.03\text{ }^{\circ}\text{C}/^{\circ}\text{C}$ without active temperature control to $0.0005\text{ }^{\circ}\text{C}/^{\circ}\text{C}$ with active control. We improved the temperature stability of the instrument with only the ThermoCube chiller, for which we extended the cooling loop to a heat exchanger attached to the fan of the electronics section. This way, the air going into the electronics section for ventilation is actively cooled. Moreover, we have put the analyzer in an enclosed box to add an extra layer of temperature isolation. With these modifications the temperature variability of T_{electr} improved to $0.21\text{ }^{\circ}\text{C}/^{\circ}\text{C}$, and to $0.07\text{ }^{\circ}\text{C}/^{\circ}\text{C}$ for T_{cell} .

2.6 Measurements of COS from flasks

10 Besides *in situ* measurements, flask or canister measurements can provide a valuable tool for providing information about ambient concentrations of COS as well (e.g. Montzka et al., 2004; Montzka et al., 2007; White et al., 2010; Blonquist et al., 2011). We filled flasks from cylinders so that we could analyse these flasks on the QCLS and the GC-MS at NOAA/ESRL (Montzka et al., 2007). This allowed us to assess the accuracy of the flask measurements and check the methodology for assigning values to our calibration cylinders. Besides testing the QCLS for continuous *in situ* measurements we explore the suitability of the QCLS for measuring air from flasks. We developed a means to analyze flasks; this allowed a comparison between the QCLS and GC-MS measurements from the NOAA Boulder laboratory (Montzka et al., 2007). For this analysis we filled flasks from cylinders, such that, besides the instrument comparison, we were able to assess the accuracy of the flask measurements and check the methodology for assigning values to our working standards based on the NOAA/ESRL calibration standards. Four pairs of glass flasks with a volume of 2.5 L were filled to 2.5 bar with dry air from two of our working standards and the two NOAA/ESRL calibration standards. The working standards were calibrated for COS by comparing their response on the QCLS to the NOAA/ESRL calibration standards as described in Sect. 2.2.2. The four cylinders contained COS concentrations between 447.8 and 486.6 ppt. The flask measurement setup using the QCLS is similar to that of routine tower measurements but with a few modifications. A pressure sensor was added to monitor the pressure in the flasks. Furthermore, a diaphragm pump with a shut-off valve was used to remove residual air from previous flask measurements in the connecting tube, and to test the lines for leaks. In Fig. 5 the placement of the pressure sensor and diaphragm pump is indicated; they are separated from the field measurement setup as it is used for flask measurements only. Air from reference and target tanks was introduced before measuring the actual flask sample to calibrate the measurements to the NOAA scale. We did not use measurements of a reference gas after the flask measurement as the stability of the measurements is affected by the larger pressure difference between the flask and the reference cylinder. Two measurements of 2 minutes each were done on each flask and the results were averaged to derive the final value. The results of these flask measurements will be shown in Sect. 3.2.

We also measured dry air flask samples to test if the water vapor correction that we determined in Sect. 2.3 sufficiently removes the effect of water vapor on calculated mole fractions. **To that aim**, flasks were filled to ambient pressure as part of a standard flask sampling routine at the Lutjewad station (Neubert et al., 2004). The air samples are dried with a cryogenic system prior to collecting. The flasks were stored for maximum 1.5 months before being measured. The same measurement strategy as for the NOAA/ESRL comparison was used, and for these flasks two measurements of ~1.5 minutes could be done before an inlet pressure of 0.3 bar was reached. We did not observe any dependence of measured dry air mole fractions in air from the flasks with the inlet pressure below ambient. The measurement results will be shown and compared with the *in situ* QCLS measurements in Sect. 3.2.

3. Results and Discussion

10 3.1 Precision and accuracy

We assessed the measurement uncertainty and accuracy with the hourly measurements of the reference and the target gases over the period from August 2014 until April 2015. As mentioned previously, each reference and target gas was measured for two minutes. The mean value of the hourly instrument-reported and uncorrected two-minute measurements are shown in Fig. 6. In Fig. 7 the standard deviation of these measurements are shown. Figure 6 also includes the electronics temperature of the QCLS. The cylinder measurements show that concentrations can drift substantially, i.e. COS concentrations easily vary by 50 to 100 ppt. However, **on the long term**, concentration changes are not correlated with temperature, which changed by 13 K throughout the year. The concentration shift on January 7, 2015 (especially visible in CO₂) happened after eliminating the leaking solenoid valve that caused mixing of nitrogen into the tubing that delivers the reference and target gases. In October 2014 the span between the two cylinder measurements changed, which is again mostly visible in CO₂. The reason for this change is that the regulator pressure of one of the two cylinders was slightly changed, which affected the amount of dilution of nitrogen into the sample line. Although it is known that COS mole fractions can drift in cylinders over time we did not find indications that the mole fractions drifted within the measurement period. **Later in this section we discuss the cylinder drift in more detail.**

25 Figure 7 shows the two-minute standard deviations of the hourly reference gas measurements between August 2014 and April 2015. It is clear from Fig. 7 that the instrument precision cannot be captured with one single value due to its variation. **For the period marked in red the room temperature was characterized by rapid changes caused by an air conditioner. This was also a period where the instrument precision was adversely affected, indicating that temperature stability influences the instrument precision.** In the right plot of Fig. 7 the histograms of the standard deviations are shown for the periods before and after improvement of the temperature stability with the black/gray colors corresponding to that in the left figure. The data show that for COS and CO₂ the period after improving the temperature stability has an improved mean standard deviation compared to the period before the temperature improvement (from 6.6 to 4.8 ppt for COS and from 0.14 to 0.06

ppm for CO₂). However, as there were also moments with better precision in August and September 2014, there is no consistent relation between temperature stability and instrument precision. What we do see is that the instrument precision changed every time that the mirror alignment was changed, indicated by blue vertical lines. During periods that the mirror alignment was not changed (January 2015 and April 2015) the instrument precision was stable. These results show that the instrument precision can largely be affected by the mirror alignment.

The overall uncertainty of the measurements consists of uncertainties associated with the scale transfer, water vapor corrections and the measurement precision. Table 5 summarizes the different uncertainty contributions as well as the overall uncertainty for measurements of COS, CO₂ and CO. The overall uncertainty is calculated as the quadrature sum (square root of the sum of squares) of the individual uncertainties. The overall uncertainty that is presented in Table 5 characterizes the typical measurement uncertainty but may vary depending on the conditions. For example, the measurement precision varies, as we discussed in the previous paragraph, and the water vapor correction uncertainty scales with the COS, CO₂ and CO concentrations. Furthermore, in Table 5 and Fig. 7 we present one-second precisions, but also precisions on longer timescales are relevant to reflect the influence of instrumental drift. Typically, precision depends on the length of the averaging period, where it can either decrease for Gaussian noise or increase for instrument drift. We will see in Fig. 8 (discussed later in this section) that for two-minute averaged target measurements the standard deviation varies between 4.9 and 6.6 ppt. This reflects the typically higher values for precision on longer timescales than one second. We provide the numbers for COS here, but the same holds for CO₂ and CO.

We have now presented the total uncertainty for long-term concentration monitoring at an atmospheric station, but not all uncertainties are relevant for every type of analyses. If the data are to be compared across different sites, then the data should be on the same scale and the accuracy of the measurements is important. However, if data from the same site and the same instrumentation are compared for example to do flux-gradient analysis, then there is a less stringent need for accuracy and the short-term precision is more important. If the uncertainties related to the transfer of the scales are not taken into account then the total uncertainty would be 5.2 ppt for COS, 0.11 ppm for CO₂ and 1.4 ppb for CO. This includes both the uncertainty of the water vapor correction and the one-second measurement precision. The uncertainty of the water vapor correction cannot simply be ignored as gradients in H₂O can exist as well. We observed that COS concentrations between 40 and 7 m differ on average by 1.7 ± 5.3 ppt during daytime and 11.4 ± 33.0 ppt during nighttime. Given that the gradients should be larger than the uncertainties mentioned above, the daytime gradients are too small to be able to use the flux-gradient method. The suitability of the flux-gradient method is further dependent on the choice of the measurement height (gradients are larger closer to the surface), the size of the fluxes at a given site and the turbulence conditions. Even though the use of the flux-gradient method is limited by the measurement uncertainty that we found here, COS profile measurements from this instrument can still be useful to derive storage fluxes, as nighttime gradients are typically larger.

Additional to the uncertainties presented in Table 5 we have observed that COS can decrease over time in uncoated aluminum cylinders. First, we did not find indications that our working standards drifted during field measurements at the Lutjewad station; calibrations in July 2014 and March 2015 showed a decrease of only 2.2 and 1.1 ppt for cylinder #1 and #2 over this 8-month period, which is well within the measurement uncertainty. However, a re-calibration in November 2015 showed a decrease of 18.2 and 24.1 ppt in these cylinders, a decrease with a rate of 2.3 and 3 ppt per month, while cylinder #3 only changed by 1.9 ppt. Cylinder #3 is not different from #1 and #2 (they are all uncoated aluminum cylinders), but cylinder #1 and #2 were stored with a pressure of 25 and 40 bar, which is lower than 130 bar for cylinder #3. We cannot confirm if the drift in the cylinders is related with the cylinder pressure; however, it is the only difference that we were able to find. Experience with cylinders over the past 15 years at NOAA indicates that COS in Aculife treated aluminum cylinders is typically much more stable than untreated aluminum cylinders. A potential method to improve COS calibrations is to calibrate these using a ppb-level standard accurately diluted to a range of desired COS concentrations (LaFranchi et al., 2015), provided the ppb-level standard is less susceptible to drifting COS concentrations. Applying this method could improve the accuracy of the calibration if the COS concentrations can be accurately and precisely provided over a broad range, thereby allowing for a more accurate determination of instrument response and a calibration curve. Besides that, this method will aid in assessing the stability of working standards.

Our COS measurements are reported on the NOAA-2004 scale, and can be compared to the observations from the global network of NOAA/ESRL (e.g. Montzka et al., 2007). The same holds for CO₂ and CO on the WMO-X2007 CO₂ scale and WMO-X2004 CO scale. To test the accuracy of the measurements against the NOAA or WMO scale we analyzed the measurements of one/two target standards after application of two corrections: (1) a correction factor as obtained from response curves of CO₂ and CO to transfer the data to the WMO scale and (2) a bias correction using a reference standard to remove instrument drift and to calibrate the data with the NOAA scale for COS. For measurements in March and April 2015 we determined hourly response curves from measurements of two working standards. For every species we took the standards with the outer concentration values such that the response curves span a wide concentration range. To analyze the need to determine hourly response curves, we corrected the data in two ways: (1) with a single bias correction for COS and a fixed response curve for CO₂ and CO, i.e. the one that we determined in the laboratory (see Fig. 2), shown in the left plots of Fig. 8, and (2) with changing response curves determined from the hourly working standard measurements, shown in the right plots of Fig. 8. For the fixed response curve, measurements of two target standards are shown, for the hourly response curve only one is shown, as two of the three working standards were needed to determine the response curve. After the corrections are applied, the mean offset of the measurements is within 3.3 ± 0.2 ppt for COS, 0.05 ± 0.003 ppm for CO₂ and 1.7 ± 0.1 ppb for CO over the period of 35 days. The standard errors indicate that the mean offset is significantly different from 0. Still, the offsets are within the expected uncertainty based on the relevant uncertainties listed in Table 5, namely the transfer to working standards and measurement calibration of which the quadrature sum is 4.0 ppt COS, 0.17 ppm CO₂ and 2.4 ppb CO. The fact that the mean offsets are on the same order of magnitude as the expected uncertainties indicates that

these uncertainties listed in Table 5 properly envelope the uncertainties of the field measurements. The slightly larger deviation from the CO assigned value for one of the two working standards with the fixed response correction is because this standard has a higher concentration than the reference standard (237.2 against 97.8 ppb) and therefore the uncertainty due to the bias correction is larger. The other target standard, which has a concentration of 119.1 ppb, is closer to the reference standard and only shows a deviation of -0.6 ppb. For CO₂ it is visible that using the fixed response curve gives a bias up to 0.2 ppm in the target measurements, which is not visible when using the hourly response curve. This bias is an effect of the fact that the response did change in the first 10 days of the period, after which it became stable. We could not relate the changing response curve to any parameter such as temperature or pressure. We did notice that the response curve was changing only in the period after the instrument was transported. A potential reason for the change could therefore be that the instrument still had to be stabilized after transportation. We did not find indications that the response curve for CO₂ changed outside the period between March and April. Except for the fact that the hourly response curve corrects for the changing response curve for CO₂, the changing response curve does not significantly remove scatter compared to the fixed response curve for CO₂ and CO; that is, the standard deviation of target measurements for the fixed response curve are not substantially lower when the hourly response curve was applied. Also the target measurements are not consistently closer to the assigned values when the hourly response curve was applied. Furthermore, the fact that the use of hourly response curves does not give lower standard deviations for target measurements of COS compared to when the single bias correction is used, indicates again that the response curves cannot accurately be determined for this species, which we discussed in Sect. 2.2.1 and 2.2.2 as well. As we have not seen indications that the response curve for CO₂ changes outside the period between March and April, we do not see the need to frequently determine the response curves with multiple standards. Moreover, if the response would change outside of the period in March and April as well, then the effect only reaches up to 0.2 ppm for CO₂. Taking into account logistical reasons (use of cylinder gases) we suggest correcting data with a single bias correction for COS, and with a fixed response curve for CO₂ and CO as determined once with NOAA/ESRL calibration standards (see Sect. 2.2.1), together with a single bias correction from a reference standard.

3.2 Measurement comparison

In this section we make different measurement comparisons. First, we compare COS flask measurements using the QCLS with those analyzed with the GC-MS at NOAA/ESRL. Second, we compare QCLS *in situ* and flask measurements of COS in order to test if the water vapor correction that we determined in Sect. 2.3 sufficiently removes the effect of water vapor on calculated mole fractions. Third, we compare the *in situ* CO₂, CO and H₂O measurements from the QCLS with those of a collocated CRDS to assess the calibration strategy and water vapor correction presented in this study. Based on the previous sections we applied the following corrections to the *in situ* QCLS data: (1) Before March 25, 2015, the TDLWINTTEL water vapor correction was applied with broadening coefficients 1.5 for COS and CO and 1.7 for CO₂. On top of this correction we applied a linear water correction curve for COS, CO₂ and CO as obtained with the TDLWINTTEL correction turned *on*. After March 25, 2015, the TDLWINTTEL correction was turned off and we applied the correction curve from Fig. 4 as obtained

with the TDLWINTTEL correction *off*; (2) the calibration correction curves as obtained in Sect. 2.2.1, Fig. 2, were applied to transfer the data to the WMO scales for CO₂ and CO; (3) a bias correction was applied to remove instrument drift and to calibrate the data with the NOAA scale for COS, here we used the same, and single, reference standard over the whole measurement period.

5

For the QCLS and GC-MS instrument comparison an overview of the measurements is given in Fig. 9 where the flask pair measurements are averaged and are shown as the deviation from the assigned cylinder value (see also Sect. 2.6). The precision within the flask pair is shown by the error bars. The comparison demonstrates the uncertainties associated with the transfer to the NOAA scale (see Table 5). For the first measurement at the QCLS (orange) three of the four flask pairs are within 1.5 ppt of the assigned value and one flask pair deviates by 7 ppt. This is similar to the GC-MS measurements at NOAA where one of the four flask pairs deviates further from the assigned values. However, this holds for another flask pair than at the QCLS. For the second QCLS measurement (blue) one flask pair has drifted on average by 12 ppt, where other flasks remained stable within 2.5 ppt. It is unclear why the two flasks have drifted as all flasks were filled, measured and stored in the same way; however, we have kept the pair to monitor the potential drift in the future and to find out what has caused the drift. Moreover, the consistency of the flask measurements between April 2015 and January 2016 is depending on the stability of the reference standard which was used for all flask measurements. Although we observed that COS can drift in cylinders we did not find indications that the particular reference standard used for this analysis drifted over the 9 month period. **The flask pairs have a mean deviation from the assigned cylinder values of +3.0 ppt for the QCLS (1.7 ppt when excluding the drifting flask pair) and -3.0 ppt for the GC-MS measurements. Although these deviations are within the measurement uncertainty, the GC-MS measurements are consistently lower than those from the QCLS, with an average difference of 5.1 ppt (excluding the drifting flask pair). We were not able to find an explanation for this bias.**

10
15
20

The next comparison in Fig. 10. considers the *in situ* COS measurements between March 25 and April 29, 2015 and 11 dry air flask samples measured by the same QCLS. ~~The time delay related to the transit of air in the inlet between sampling of air for flasks and QCLS measurements is assumed to be the same, as both sampling systems have a flow rate of 2 L min⁻¹ and the sample tubing of the systems has the same size.~~ The flasks are flushed for an hour before closing, but because of mixing in the flask we assume that the flask sample represents the last 15 minutes, therefore we average the *in situ* measurements over these 15 minutes. The flask samples have an inlet at 60 m height, but because the *in situ* 60 m measurements only cover the period between 9 and 15 minutes before flask closure we also include 40 m measurements to cover the last 9 minutes before the flask is closed. The average difference of COS mole fractions between the 40 and 60 m level is 0.7 ± 9.7 ppt (40 – 60 m) over the measurement period in March and April. **Although the mean difference of 0.7 ppt is not significant, the 9.7 ppt standard deviation points to actual differences between the 40 and 60 m concentrations. For the moments where the flask and *in situ* measurements are compared the difference is -4.2 ± 3.4 ppt. We neglected one flask – *in situ* pair for which the 40 and 60 m difference was larger than 10 ppt. For the remaining data we do not expect a large bias associated with including**

25
30

ongoing results from the 40 m height in the averages. Peaks and valleys in COS mole fractions shown in Fig. 10 are well covered by both sampling techniques (flasks and *in situ*); for example, the peak up to 620 ppt at April 11 is clearly visible in both the *in situ* and flask measurements. The average difference between the *in situ* and flask measurements (*in situ* – flask) is -9.7 ± 4.6 ppt. Besides neglecting one flask measurement where the 40 - 60 m difference was larger than 10 ppt, we neglected one flask sample where the *in situ* COS measurement over 15 minutes showed large variation (standard deviation of 17.5 ppt where the average standard deviation of the other periods is only 4.7 ppt) and thereby introduced an error in the comparison. The average difference of -9.7 ppt could be caused by the uncertainty of the flask and the *in situ* measurements (where the latter includes the uncertainty of the water vapor correction), the sampling bias of both *in situ* and flask measurements, as well as the storage effect of air in the flasks. In Table 5 we quantified the uncertainty related to the transfer of the scale to working standards and measurements, the water correction, and the measurement precision. Combining these uncertainties gives an uncertainty of 5.8 ppt for the flasks and 6.5 ppt for the *in situ* measurements (with the uncertainty of the water vapor correction included). The combined uncertainty of both methods is 8.7 ppt, which is only slightly smaller than the -9.7 ppt difference that we found. We were not able to test the effect of biases resulting from the sampling of flasks. We find that after applying the water correction to the *in situ* data, the flask – *in situ* difference does not show a correlation with H₂O ($R^2 = 0.01$; slope = 3.9 ppt COS per % H₂O), whereas without water vapor correction there is a correlation with H₂O ($R^2 = 0.36$; slope = 26.3 ppt COS per % H₂O), which demonstrates that our laboratory derived water vapor correction properly corrects for the effect of water vapor.

Last, we compare the minute averaged QCLS measurements for CO₂, CO, and H₂O with those made by a collocated CRDS in Fig. 11. The air samples were taken from the same height, but through a different inlet. The CRDS measurements of CO₂, and CO were performed on humid air, and were corrected for water vapor dilution and interference effects based on a set of instrument-specific correction factors determined in the laboratory of the Center for Isotope Research before field deployment (Chen et al., 2010; 2013). The comparison was only made from January 7 onwards because before that period there was the problem of nitrogen leaking into the sample air for which we used the CRDS data to correct for the dilution. There are no temporal patterns visible in the difference plot on the right of Fig. 11. The mean differences (QCLS – CRDS) are -0.12 ± 0.77 ppm for CO₂, -0.9 ± 3.8 ppb for CO and -0.01 ± 0.09 % for H₂O. For H₂O there were problems with water vapor in the sample lines in March and April, therefore we only calculated the difference for H₂O over the month January. A correlation of the CRDS and QCLS data of both CO₂ and CO (slope = 0.98) indicates that there is no concentration dependent offset. Also no correlation was found between water vapor and the difference between CRDS and the QCLS CO₂ and CO data (CO₂: slope = 0.16, $R^2 = 0.003$; CO: slope = 0.48, $R^2 = 1.9 \cdot 10^{-5}$). These results give confidence in the calibration strategy and water vapor correction presented in this study. For the period up to January 29, the TDLWINTTEL water vapor correction was turned on, and on top of this correction we applied the linear water correction curves as obtained with the TDLWINTTEL correction turned on. If we would not apply this linear correction curve on top of the TDLWINTTEL correction, then the median difference between the QCLS and CRDS measurements for this period would be -0.59 ppm for

CO₂ and -1.1 ppb for CO (against -0.02 ppm and -0.9 ppb for specifically this period when the linear correction curve is applied).

3.3 Continuous COS, CO₂, CO and H₂O observations from Lutjewad

The COS, CO₂, CO and H₂O data record obtained at the 60 m level of the Lutjewad tower is presented in Fig. 12. **The data corrections used for these data were summarized in the previous section. Besides of these corrections** we corrected for the dilution of nitrogen due to a leaking solenoid valve before January 7, 2015. We determined a dilution factor by comparing CO₂ measurements from the QCLS and a CRDS from the same location and height, under the assumption that without dilution the two analyzers measure the same concentrations (see Sect. 3.2). The percentage of dilution was calculated for CO₂ and was typically between 0.4 and 4.9 %. These dilution factors were then applied to all species.

10

The location of the Lutjewad station along the coast of the province of Groningen in the Netherlands allows the measurement of marine background air during northerly winds, and continental air during southerly winds (Van der Laan et al., 2009). Daytime CO₂ concentrations are typically correlated with elevated CO concentrations, indicating the influence of local and regional fossil fuel emissions (Van der Laan et al., 2010). Even though the data do not cover a full year cycle, it can be seen that the seasonal amplitude of CO₂ mole fractions is approximately 15 ppm with a minimum around the end of August. The seasonal variation of CO₂ for the Lutjewad measurement station is analyzed in detail by Van der Laan-Luijkx et al. (2010) and Van Leeuwen (2015). **The seasonal cycle of COS is captured well by the QCLS** and the peak-to-peak amplitude is estimated to be 96 ppt based on the two-harmonic fit. Kettle et al. (2002) showed that vegetative uptake is the flux with the largest seasonal cycle on the NH, and Montzka et al. (2007) showed that the seasonal amplitude of COS depends on the degree to which the sampled air is influenced by terrestrial ecosystems. It is therefore likely that the seasonal variation of COS observed at the Lutjewad site is influenced by vegetative uptake. In Fig. 13 we compare COS mole fractions from Lutjewad with that from three other sites as measured from flask samples with a GC-MS by NOAA/ESRL (Montzka et al., 2007). The flask samples cover data between 2000 and 2015 for Wisconsin, United States (LEF) and Mauna Loa, United States (MLO) and between 2001 and 2015 for Mace Head, Ireland (MHD). These data are an update of those presented in Montzka et al. (2007) (data available at: ftp://ftp.cmdl.noaa.gov/hats/carbonyl_sulfide/). The flask measurements are plotted as function of time of the year. The high altitude MLO site is less directly influenced by terrestrial ecosystems and therefore shows only small seasonal variation, in contrast to the LEF site, which is largely influenced by (forested) continental air. The Lutjewad COS mole fractions are most consistent with measurements from MHD, which can be expected since both stations are coastal sites and are located at similar latitudes. The seasonal amplitude of COS at MHD and Lutjewad is in between that of LEF and MLO, most likely because both sites are not solely influenced by marine or continental air but by both types of air masses. The COS mole fraction has a minimum in September and October, and is a few weeks later than the minimum of the CO₂ mole fraction. Montzka et al. (2007) and Blonquist et al. (2011) also observed a COS minimum later than that of CO₂. They reasoned that this difference is due to the fact that at the end of the growing

20
25
30

season COS mole fractions keep decreasing due to vegetative uptake without at the same time having a source of COS, whereas during this time of year respiration is beginning to offset assimilation in determining the ambient CO₂ mole fractions.

4. Conclusions

5 In this study we have tested a QCLS for its suitability for making accurate and high precision measurements of COS, CO₂, CO and H₂O. First, the instrument response was characterized using calibration standards and to transfer raw data to the NOAA or WMO scale. Unfortunately, the range of mole fractions in the calibration standards did not allow the COS response to be accurately determined over the entire range of measured mole fractions. Based on an analysis of different calibration methods, however, we concluded that the raw measurements and working standards were best calibrated using a
10 single bias correction for COS. From hourly paired measurements of working standards we observed changes in the response curve for CO₂ over a period of 10 days after transporting the instrument to the measurement site. However, as we have not seen indications that the response curve for CO₂ changed outside this period, and also taking into account logistical reasons (use of cylinder gases) we suggest calibrating field data with a fixed response curve for CO₂ and CO as determined once with calibration standards. Second, we investigated the needed frequency of background and reference measurements. Based on
15 laboratory tests we have shown that background measurements every six hours with reference measurements every 30 minutes (for removal of instrument drift) are sufficient to keep the standard deviation of working standard measurements within 4.6 ppt for COS, 0.08 ppm for CO₂ and 0.34 ppb for CO over a period of 19 hours. We characterized the water vapor dependence of COS, CO₂ and CO from laboratory experiments. Based on an assessment of the TDLWINTTEL water correction we determined optimal broadening coefficients for the use of the water correction within TDLWINTTEL. Besides
20 that, we presented an alternative water correction based on linear dependence of the wet air mole fractions with H₂O concentration. Furthermore, we demonstrate that a small H₂O peak close to the COS peak has caused a water vapor dependent concentration error that is larger than the direct water vapor dilution effect. This water vapor interference can be minimized by careful adjustments to the software fitting parameters and was virtually eliminated with corrections as demonstrated in Fig. 4.

25

The QCLS was set up for continuous *in situ* measurements at different heights at the tower of the Lutjewad monitoring station. Hourly target measurements were used to assess the accuracy and precision of the measurements. After application of a calibration response curve for CO₂ and CO, and a single bias correction for removal of instrument drift and to calibrate the COS measurements to the NOAA scale, the target measurements showed a mean difference with the assigned cylinder
30 value within 3.3 ppt for COS, 0.05 ppm for CO₂ and 1.7 ppb for CO over a period of 35 days. One-second precisions during reference gas flow were typically 4.3 ppt for COS, 0.04 ppm for CO₂ and 0.9 ppb for CO, however, substantial variations in the instrument precision were observed during the 7 month field campaign. We improved the temperature stability of the

QCLS by applying an additional insulation layer that is controlled by the same thermoelectric chiller as the one used for cooling the laser and detector. However, improvement of the temperature stability of the instrument did not show a consistent relation with instrument precision. We showed that variations in the precision were largely driven by mirror alignment. The different uncertainty contributions for measurements of COS, CO₂ and CO were summarized and the overall uncertainty was determined to be 6.9 ppt for COS, 0.21 ppm for CO₂ and 3.4 ppb for CO. We discussed the relevance of different uncertainty contributions for different types of analyses (e.g. differences across sites and interpreting profile measurements). Our setup with the QCLS provides sufficient accuracy and precision to detect gradients larger than 5.2 ppt for COS, 0.11 ppm for CO₂ and 1.4 ppb for CO. For the current setup in Lutjewad, the daytime gradients were too small compared to the measurement uncertainty to be able to use the flux-gradient method. QCLS measurements were compared with independent CRDS measurements for CO₂ and CO, and with dry-air flask samples at the QCLS for COS, which showed a mean difference of -9.7 ± 4.6 ppt for COS, 0.12 ± 0.77 ppm for CO₂, -0.9 ± 3.8 ppb for CO and -0.01 ± 0.09 % for H₂O. After correction for water vapor there was no correlation between the offsets and water vapor for all species. The measurement record over the 7 month period was presented and compared with NOAA/ESRL flask measurements for COS at other sites in the northern hemisphere. The peak-to-peak amplitude of COS in ambient air at the Lutjewad monitoring station was estimated to be 97 ppt, which is comparable to other coastal sites at similar latitudes in the Northern Hemisphere.

Acknowledgements

We would like to thank B.A.M. Kers, M. de Vries, H.G. Jansen and H.A. Been for their help in preparing the system for installation in the field and for maintenance of the instrumentation in Lutjewad. We thank D. Paul for the valuable discussions and suggestions, H.A. Scheeren for providing the CO₂ and CO calibrations of our working standards and J.J. Spriensma for her help in sorting out flask samples. We also acknowledge the preparation of gravimetric standards and working standards at NOAA by B. Hall, and the technical assistance of C. Siso. Finally, we would like to thank the reviewers for their valuable comments and suggestions. This work was supported by the Dutch Sector Plan Physics.

References

- Asaf, D., Rotenberg, E., Tatarinov, F., Dicken, U., Montzka, S. A., and Yakir, D.: Ecosystem photosynthesis inferred from measurements of carbonyl sulphide flux, *Nat. Geosci.*, 6, 186-190, doi: 10.1038/ngeo1730, 2013.
- Belviso, S., Schmidt, M., Yver, C., Ramonet, M., Gros, V., and Launois, T.: Strong similarities between night-time deposition velocities of carbonyl sulphide and molecular hydrogen inferred from semi-continuous atmospheric observations in Gif-sur-Yvette, Paris region, *Tellus B*, 65, 2013.

- Berkelhammer, M., Asaf, D., Still, C., Montzka, S., Noone, D., Gupta, M., Provencal, R., Chen, H., and Yakir, D.: Constraining surface carbon fluxes using in situ measurements of carbonyl sulfide and carbon dioxide, *Global Biogeochem. Cy.*, 28, 161-179, doi: 10.1002/2013GB004644, 2014.
- Berry, J., Wolf, A., Campbell, J. E., Baker, I., Blake, N., Blake, D., Denning, A. S., Kawa, S. R., Montzka, S. A., Seibt, U.,
5 Stimler, K., Yakir, D., and Zhu, Z.: A coupled model of the global cycles of carbonyl sulfide and CO₂: A possible new window on the carbon cycle, *J. Geophys. Res. Biogeo.*, 118, 842-852, doi: 10.1002/jgrg.20068, 2013.
- Blonquist, J. M., Montzka, S. A., Munger, J. W., Yakir, D., Desai, A. R., Dragoni, D., Griffis, T. J., Monson, R. K., Scott, R. L., and Bowling, D. R.: The potential of carbonyl sulfide as a proxy for gross primary production at flux tower sites, *J. Geophys. Res. Biogeo.*, 116, doi: 10.1029/2011JG001723, 2011.
- 10 Campbell, J. E., Carmichael, G. R., Chai, T., Mena-Carrasco, M., Tang, Y., Blake, D. R., Blake, N. J., Vay, S. A., Collatz, G. J., Baker, I., Berry, J. A., Montzka, S. A., Sweeney, C., Schnoor, J. L., and Stanier, C. O.: Photosynthetic Control of Atmospheric Carbonyl Sulfide During the Growing Season, *Science*, 322, 1085-1088, doi: 10.1126/science.1164015, 2008.
- Chen, H., Winderlich, J., Gerbig, C., Hofer, A., Rella, C. W., Crosson, E. R., Van Pelt, A. D., Steinbach, J., Kolle, O.,
15 Beck, V., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Santoni, G. W., and Wofsy, S. C.: High-accuracy continuous airborne measurements of greenhouse gases (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique, *Atmos. Meas. Tech.*, 3, 375-386, doi: 10.5194/amt-3-375-2010, 2010.
- Chen, H., Karion, A., Rella, C. W., Winderlich, J., Gerbig, C., Filges, A., Newberger, T., Sweeney, C., and Tans, P. P.: Accurate measurements of carbon monoxide in humid air using the cavity ring-down spectroscopy (CRDS)
20 technique, *Atmos. Meas. Tech.*, 6, 1031-1040, doi: 10.5194/amt-6-1031-2013, 2013.
- Commane, R., Herndon, S. C., Zahniser, M. S., Lerner, B. M., McManus, J. B., Munger, J. W., Nelson, D. D., and Wofsy, S. C.: Carbonyl sulfide in the planetary boundary layer: Coastal and continental influences, *J. Geophys. Res. Atmos.*, 118, 8001-8009, doi: 10.1002/jgrd.50581, 2013.
- Commane, R., Meredith, L. K., Baker, I. T., Berry, J. A., Munger, J. W., Montzka, S. A., Templer, P. H., Juice, S. M.,
25 Zahniser, M. S., and Wofsy, S. C.: Seasonal fluxes of carbonyl sulfide in a midlatitude forest, *P. Natl. Acad. Sci. USA*, doi: 10.1073/pnas.1504131112, 2015.
- LaFranchi, B., Bambha, R., Schrader, P., and Michelsen, H.: Characterization of continuous OCS, CO and CO₂ measurements at a tower site in Livermore, CA USA, 18th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT), 13-17 September 2015, A10, 2015.
- 30 Kettle, A., Kuhn, U., von Hobe, M., Kesselmeier, J., and Andreae, M.: Global budget of atmospheric carbonyl sulfide: Temporal and spatial variations of the dominant sources and sinks, *J. Geophys. Res. Atmos.*, 107, 4658, doi: 10.1029/2002JD002187, 2002.

- Maseyk, K., Berry, J. A., Billesbach, D., Campbell, J. E., Torn, M. S., Zahniser, M., and Seibt, U.: Sources and sinks of carbonyl sulfide in an agricultural field in the Southern Great Plains, *P. Natl. Acad. Sci. USA*, 2015 112 (46) 14162-14167, doi: 10.1073/pnas.1319132111, 2014.
- McManus, J. B., Zahniser, M. S., Nelson, D. D., Jr., Shorter, J. H., Herndon, S., Wood, E., and Wehr, R.: Application of quantum cascade lasers to high-precision atmospheric trace gas measurements, *Opt. Eng.*, 49, 111124, doi: 10.1117/1.3498782, 2010.
- Meredith, L. K., Commane, R., Munger, J. W., Dunn, A., Tang, J., Wofsy, S. C., and Prinn, R. G.: Ecosystem fluxes of hydrogen: a comparison of flux-gradient methods, *Atmos. Meas. Tech.*, 7, 2787-2805, doi: 10.5194/amt-7-2787-2014, 2014.
- Miller, B. R., Weiss, R. F., Salameh, P. K., Tanhua, T., Grealley, B. R., Mühle, J., and Simmonds, P. G.: Medusa: A Sample Preconcentration and GC/MS Detector System for in Situ Measurements of Atmospheric Trace Halocarbons, Hydrocarbons, and Sulfur Compounds, *Anal. Chem.*, 80, 1536-1545, doi: 10.1021/ac702084k, 2008.
- Montzka, S., Aydin, M., Battle, M., Butler, J., Saltzman, E., Hall, B., Clarke, A., Mondeel, D., and Elkins, J.: A 350-year atmospheric history for carbonyl sulfide inferred from Antarctic firn air and air trapped in ice, *J. Geophys. Res. Atmos.*, 109, D22302, doi: 10.1029/2004JD004686, 2004.
- Montzka, S. A., Calvert, P., Hall, B. D., Elkins, J. W., Conway, T. J., Tans, P. P., and Sweeney, C.: On the global distribution, seasonality, and budget of atmospheric carbonyl sulfide (COS) and some similarities to CO₂, *J. Geophys. Res. Atmos.*, 112, D09302, doi: 10.1029/2006JD007665, 2007.
- Nelson, D.D., McManus, B., Urbanski, S., Herndon, S., and Zahniser, M.: High precision measurements of atmospheric nitrous oxide and methane using thermoelectrically cooled mid-infrared quantum cascade lasers and detectors, *Spectrochim. Acta. A.*, 60, 3325-3335, doi: 10.1016/j.saa.2004.01.033, 2004.
- Neubert, R. E. M., Spijkervet, L. L., Schut, J. K., Been, H. A., and Meijer, H. A. J.: A Computer-Controlled Continuous Air Drying and Flask Sampling System, *J. Atmos. Ocean. Tech.*, 21, 651-659, doi: 10.1175/1520-0426(2004)021<0651:ACCADA>2.0.CO;2, 2004.
- Novelli, P. C., Masarie, K. A., Lang, P. M., Hall, B. D., Myers, R. C., and Elkins, J. W.: Reanalysis of tropospheric CO trends: Effects of the 1997–1998 wildfires, *J. Geophys. Res. Atmos.*, 108, doi: 10.1029/2002JD003031, 2003.
- Protoschill-Krebs, G. and Kesselmeier, J.: Enzymatic Pathways for the Consumption of Carbonyl Sulphide (COS) by Higher Plants*, *Bot. Acta*, 105, 206-212, doi: 10.1111/j.1438-8677.1992.tb00288.x, 1992.
- ProtoschillKrebs, G., Wilhelm, C., and Kesselmeier, J.: Consumption of carbonyl sulphide (COS) by higher plant carbonic anhydrase (CA), *Atmos. Environ.*, 30, 3151-3156, doi: 10.1016/1352-2310(96)00026-X, 1996.
- Rella, C. W., Chen, H., Andrews, A. E., Filges, A., Gerbig, C., Hatakka, J., Karion, A., Miles, N. L., Richardson, S. J., Steinbacher, M., Sweeney, C., Wastine, B., and Zellweger, C.: High accuracy measurements of dry mole fractions of carbon dioxide and methane in humid air, *Atmos. Meas. Tech.*, 6, 837-860, doi: 10.5194/amt-6-837-2013, 2013.

- Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Chris Benner, D., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J. -, Gamache, R. R., Harrison, J. J., Hartmann, J. -, Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R. J., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Müller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., and Wagner, G.: The HITRAN2012 molecular spectroscopic database, *J. Quant. Spectrosc. Ra.*, 130, 4-50, doi: <http://dx.doi.org/10.1016/j.jqsrt.2013.07.002>, 2013.
- Sandoval-Soto, L., Stanimirov, M., von Hobe, M., Schmitt, V., Valdes, J., Wild, A., and Kesselmeier, J.: Global uptake of carbonyl sulfide (COS) by terrestrial vegetation: Estimates corrected by deposition velocities normalized to the uptake of carbon dioxide (CO₂), *Biogeosciences*, 2, 125-132, 2005.
- Santoni, G. W., Lee, B. H., Goodrich, J. P., Varner, R. K., Crill, P. M., McManus, J. B., Nelson, D. D., Zahniser, M. S., and Wofsy, S. C.: Mass fluxes and isofluxes of methane (CH₄) at a New Hampshire fen measured by a continuous wave quantum cascade laser spectrometer, *J. Geophys. Res. Atmos.*, 117, D10301, doi: 10.1029/2011JD016960, 2012.
- Seibt, U., Kesselmeier, J., Sandoval-Soto, L., Kuhn, U., and Berry, J. A.: A kinetic analysis of leaf uptake of COS and its relation to transpiration, photosynthesis and carbon isotope fractionation, *Biogeosciences*, 7, 333-341, doi: 10.5194/bg-7-333-2010, 2010.
- Stimler, K., Nelson, D., and Yakir, D.: High precision measurements of atmospheric concentrations and plant exchange rates of carbonyl sulfide using mid-IR quantum cascade laser, *Glob. Change Biol.*, 16, 2496-2503, doi: 10.1111/j.1365-2486.2009.02088.x, 2010a.
- Stimler, K., Montzka, S. A., Berry, J. A., Rudich, Y., and Yakir, D.: Relationships between carbonyl sulfide (COS) and CO₂ during leaf gas exchange, *New Phytol.*, 186, 869-878, doi: 10.1111/j.1469-8137.2010.03218.x, 2010b.
- van der Laan, S., Neubert, R. E. M., and Meijer, H. A. J.: Methane and nitrous oxide emissions in The Netherlands: ambient measurements support the national inventories, *Atmos. Chem. Phys.*, 9, 9369-9379, doi: 10.5194/acp-9-9369-2009, 2009.
- Van der Laan, S., Karstens, U., Neubert, R. E. M., Van der Laan-Luijkx, I. T., and Meijer, H. A. J.: Observation-based estimates of fossil fuel-derived CO₂ emissions in the Netherlands using Delta 14C, CO and 222Radon, *Tellus B*, 62, 389-402, doi: 10.1111/j.1600-0889.2010.00493.x, 2010.
- van der Laan-Luijkx, I. T., Neubert, R. E. M., van der Laan, S., and Meijer, H. A. J.: Continuous measurements of atmospheric oxygen and carbon dioxide on a North Sea gas platform, *Atmos. Meas. Tech.*, 3, 113-125, 2010.
- Van Leeuwen, C.: Highly precise atmospheric oxygen measurements as a tool to detect leaks of carbon dioxide from Carbon Capture and Storage sites, Ph.D. thesis, University of Groningen, The Netherlands, 2015.

- White, M. L., Zhou, Y., Russo, R. S., Mao, H., Talbot, R., Varner, R. K., and Sive, B. C.: Carbonyl sulfide exchange in a temperate loblolly pine forest grown under ambient and elevated CO₂, *Atmos. Chem. Phys.*, 10, 547-561, doi: 10.5194/acp-10-547-2010, 2010.
- 5 Wohlfahrt, G., Brilli, F., Hoertnagl, L., Xu, X., Bingemer, H., Hansel, A., and Loreto, F.: Carbonyl sulfide (COS) as a tracer for canopy photosynthesis, transpiration and stomatal conductance: potential and limitations, *Plant Cell Environ.*, 35, 657-667, doi: 10.1111/j.1365-3040.2011.02451.x, 2012.
- Xiang, B., Nelson, D. D., McManus, J. B., Zahniser, M. S., Wehr, R. A., and Wofsy, S. C.: Development and field testing of a rapid and ultra-stable atmospheric carbon dioxide spectrometer, *Atmos. Meas. Tech.*, 7, 4445-4453, doi: 10.5194/amt-7-4445-2014, 2014.
- 10 Zhao, C. L. and Tans, P. P.: Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air, *J. Geophys. Res. Atmos.*, 111, doi: 10.1029/2005JD006003, 2006.
- Zhao, F. and Zeng, N.: Continued increase in atmospheric CO₂ seasonal amplitude in the 21st century projected by the CMIP5 Earth system models, *Earth Syst. Dynam.*, 5, 423-439, doi: 10.5194/esd-5-423-2014, 2014.

Tables

Table 1 CO₂ and CO calibration values together with 1- σ uncertainties of the calibration curves as obtained with the QCLS and a cavity ringdown spectrometer (CRDS) for comparison of calibration results. COS calibration values could not be compared with that from other instrumentation.

	CO ₂ [ppm]		CO [ppb]	
	QCLS	CRDS	QCLS	CRDS
Cyl. #1	412.33 \pm 0.12	412.43 \pm 0.08	97.8 \pm 1.7	97.6 \pm 3.0
Cyl. #2	398.30 \pm 0.12	398.13 \pm 0.08	119.1 \pm 1.7	119.1 \pm 3.0

5

Table 2 COS calibration values for our real-air working standards as obtained from three different calibration approaches: (1) with response from the two NOAA/ESRL calibration standards and a zero-point, (2) with the two NOAA/ESRL calibration standards and the curve *not* forced through a zero-point, and (3) using a single bias correction. The NOAA/ESRL calibration standards have COS concentrations 447.8 and 486.6 ppt. The calibration measurement was repeated three times; results are shown as the average over the three measurement and uncertainties indicate the standard deviation over the three measurements.

	Cyl A	Cyl B	Cyl C	Cyl D
1. Through NOAA/ESRL standards and 0	393.0 ± 2.2	448.8 ± 4.7	473.6 ± 1.5	504.1 ± 0.7
2. Through NOAA/ESRL standards and not 0	379.2 ± 8.3	445.5 ± 4.4	474.7 ± 1.9	510.8 ± 4.7
3. Single bias correction	390.9 ± 2.6	445.8 ± 3.9	476.6 ± 2.4	506.6 ± 2.1

Table 3 Standard deviation over minute averaged data of COS, CO₂ and CO over the 19 hour measurement period for different correction frequencies using reference measurements.

	uncorrected	1 hr. corr.	30 min. corr.	15 min. corr.
COS stdev. [ppt]	11.8	9.6	5.3	4.6
CO ₂ stdev. [ppm]	0.26	0.12	0.09	0.08
CO stdev. [ppb]	0.88	0.70	0.33	0.34

Table 4 Different water vapor correction strategies based on the software fitting parameters (standard or split fit) and with the TDLWINTTEL correction turned on or off. In case the TDLWINTTEL correction is turned on the values indicate the broadening coefficient used for the different species, and in the case it is turned off the values indicate the slope of the correction curve as determined in Fig. 4 with $y = slope * H_2O + 1$ with H_2O in percent and y the wet/dry ratio of the gas.

	TDLWINTTEL correction on/off	Broadening coefficient or slope		
		COS	CO ₂	CO
Standard fit	on	-*	2.15	1.0
	off	0.030	-0.0146	-0.008
Split fit	on	1.0	2.15	1.0
	off	-0.008	-0.0149	-0.008

5

*No broadening coefficient could be derived; however, we found that for a broadening coefficient of 1.5 with the standard fit the slope of the curve for COS is equal to 0.031, which can be applied as an extra correction on top of the TDLWINTTEL correction.

10

Table 5 Uncertainty contributions and the overall uncertainty for measurements of COS (447.8-486.6 ppt), CO₂ (354 – 426 ppm) and CO (94-384 ppb), for the range of H₂O (0-2.1 %).

Uncertainty contributions	COS [ppt]	CO ₂ [ppm]	CO [ppb]
Repeatability of the NOAA or WMO scale	2.1	0.07	2.0
Transfer scale to working standards (1- σ)*	2.8	0.12	1.7
Measurement calibration **	2.8	0.12	1.7
Water vapor correction (1- σ)	2.9	0.10	1.1
Measurement precision (1-sec.)***	4.3	0.04	0.9
<i>Overall uncertainty</i>	6.9	0.21	3.4

*Average uncertainty over 4 cylinders in Table 2 (method 3).

**Using the single bias correction (see Sect. 2.2.2) it is the same as transferring the scale to the working standards.

5 ***The 50% percentile of the short-term precision of working standard measurements in March-April 2015 (Fig 7.).

Figures

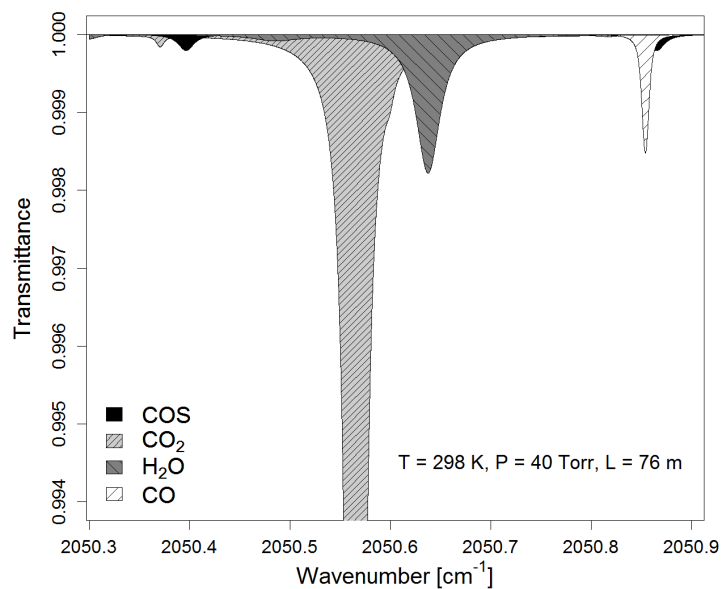


Figure 1 Simulated transmission spectrum of ambient concentrations of COS (500 ppt), CO₂ (500 ppm), H₂O (1.5 %) and
5 CO (200 ppb) with sample cell conditions: temperature 298 K, pressure 53.3 hPa (40 Torr) and the absorption path length 76
m. A small water band at 2050.5 cm⁻¹ can interfere with COS at 2050.4 cm⁻¹ and can affect the COS correction for water
vapor without a split fit at 2050.45 cm⁻¹ (Sect. 2.3).

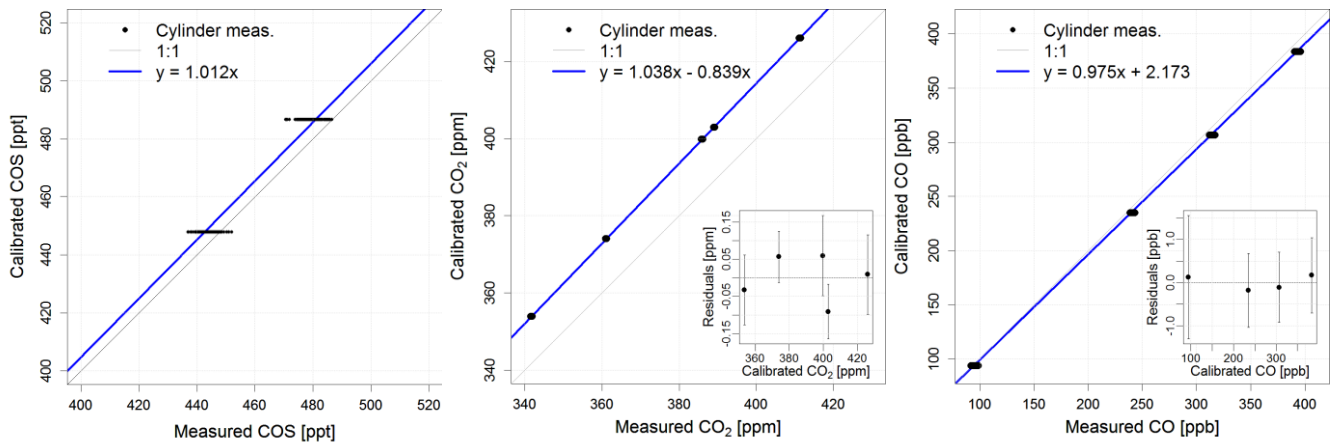


Figure 2 Response curves determined with NOAA/ESRL calibration standards including the residuals of the fit. The NOAA/ESRL calibration standards are calibrated on NOAA-2004 COS scale, WMO-X2007 CO₂ scale (Zhao and Tans, 2006) and WMO-2004 CO scale (Novelli et al., 2003). Each standard was measured for 5 minutes. The measured mole fractions shown on the x axes are mole fractions calculated through the TDLWINTTEL software but are corrected for drift using a reference standard. The drift was on the order of 8 ppt for COS, 3 ppm for CO₂ and 5 ppb for CO and would have given biased results if this was not corrected for.

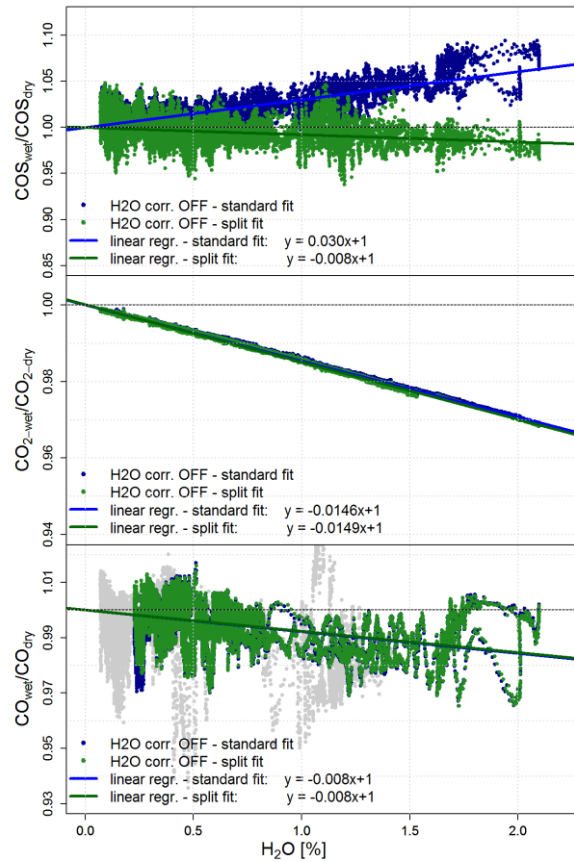
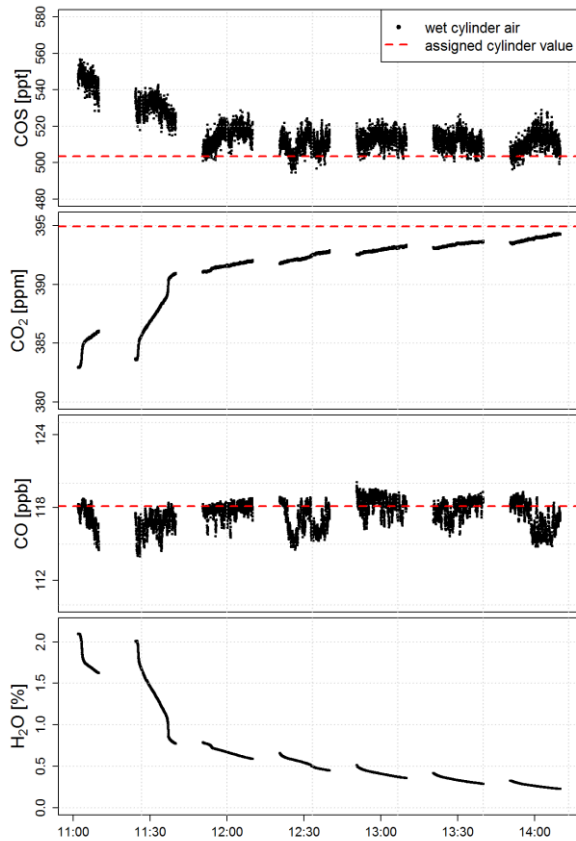


Figure 3 (left) Water vapor experiment where cylinder air was humidified with wet silica gel and no water vapour correction was applied to the data.

Figure 4 (right) Wet over dry ratio of COS, CO₂ and CO versus H₂O as obtained from multiple water vapor experiments with humidified cylinder air. The data show the linear dependence with H₂O when the TDLWINTTEL water vapor correction was turned off with the standard fit (blue) and with the split fit (green).

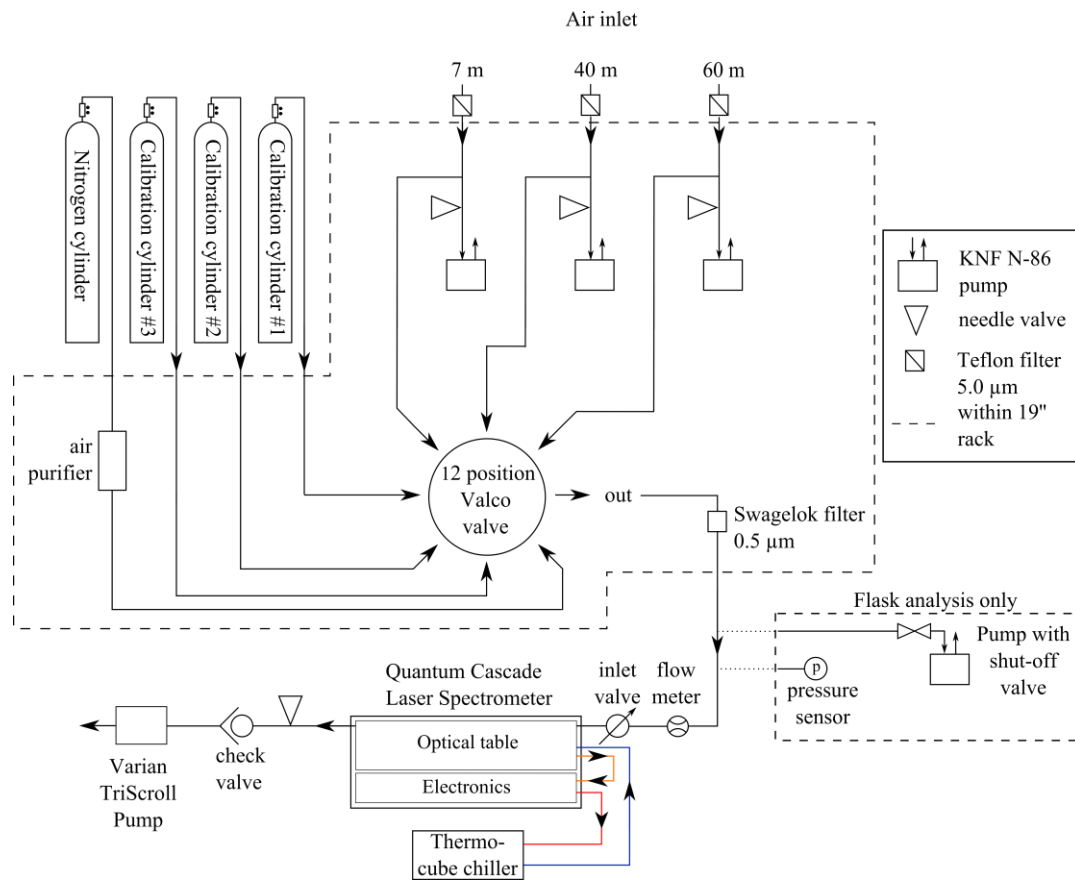


Figure 5 Schematic overview of the instrument setup for tower profile measurements at the Lutjewad monitoring station.

5 The pressure sensor and pump with shut-off valve were added for flask measurements only (see Sect. 2.6).

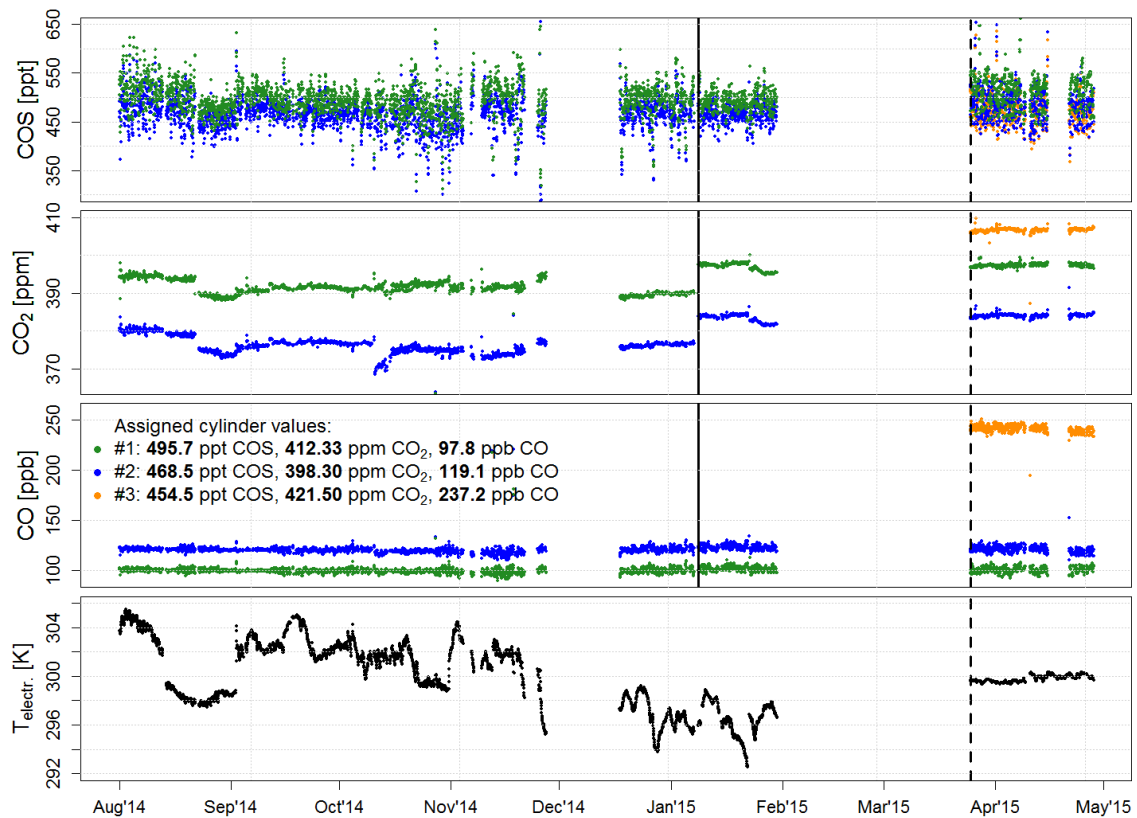


Figure 6 Mean concentrations of hourly measurements of working standards conducted while the instrument was in the field from August 2014 until April 2015 together with electronics temperature. **The data shown here are uncorrected data and are not calibrated with a response curve. The concentrations are therefore not necessarily close to the assigned cylinder values.**

- 5 The solid vertical line at January 7, 2015 indicates the moment where we changed the setup with a solenoid valve to a Valco valve for switching to nitrogen, as the solenoid valve was found leaking. The dashed vertical line at March 25, 2015 indicates the moment where we improved the temperature stability by actively cooling the electronics section and putting the analyzer in an enclosed box, which resulted in substantially smaller temperature fluctuations than before. From this moment onwards also an extra working standard was measured every hour to ascertain if the instrument response was stable over a
- 10 period of 35 days (orange). The gap in the data record in December 2014 and February-March 2015 is because of tests with the QCLS in the laboratory.

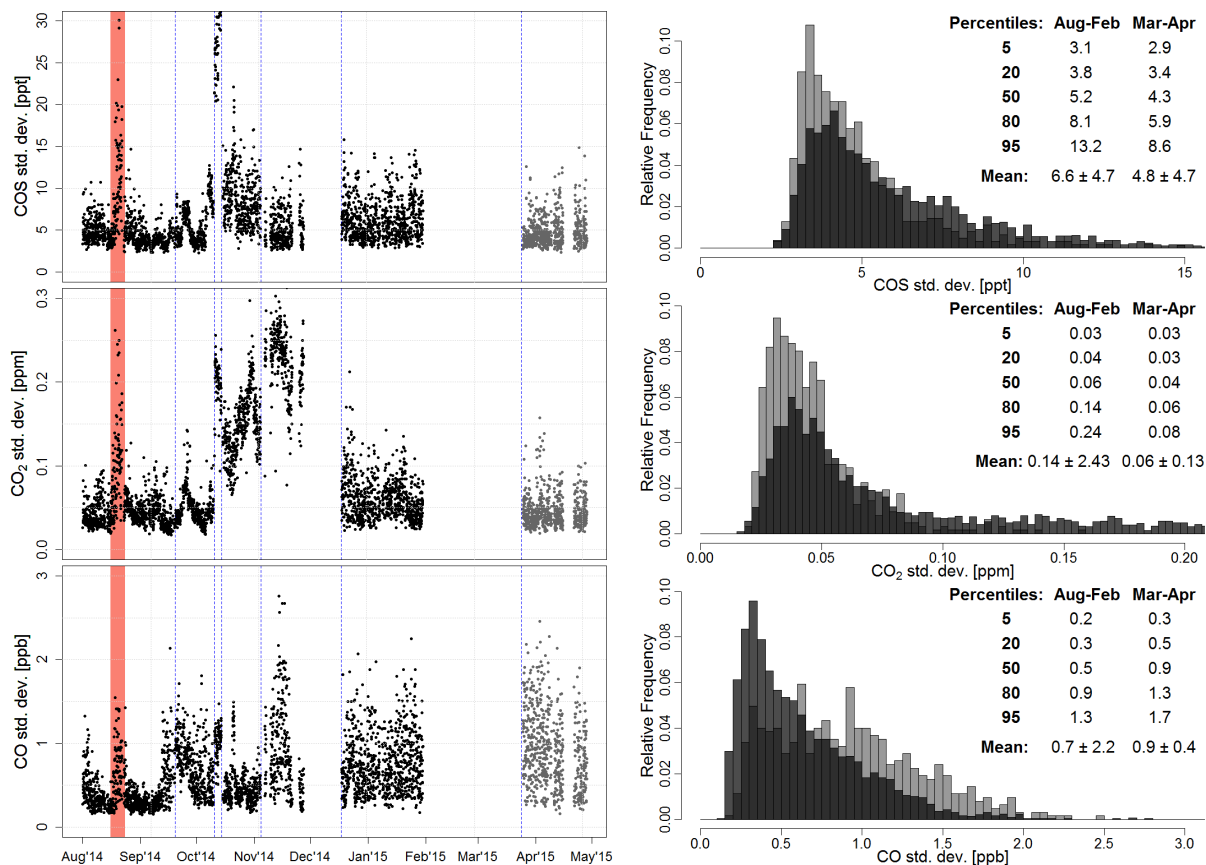


Figure 7 Left: Standard deviation over two minutes of hourly measurements of one of the working standards as measured from August 2014 until April 2015. The gray (black) colored data indicate the period after (before) modifications were made to improve the temperature stability (such that T_{electr} varied with $0.21\text{ }^{\circ}\text{C}$ with every degree of changing T_{room} , against $0.91\text{ }^{\circ}\text{C}/^{\circ}\text{C}$ before modifications, see Section 2.5). For the period that is marked with a red background the room temperature was characterized by rapid changes caused by an air conditioner. Blue vertical (dashed) lines indicate moments where the mirror alignment was changed. Right: Histogram of standard deviations as shown in the left figure with the black/gray colors corresponding to the colors in the left figure. Note that the dark histogram is transparent. Also included are the mean standard deviations as well as an overview of percentiles because the data do not show a Gaussian distribution.

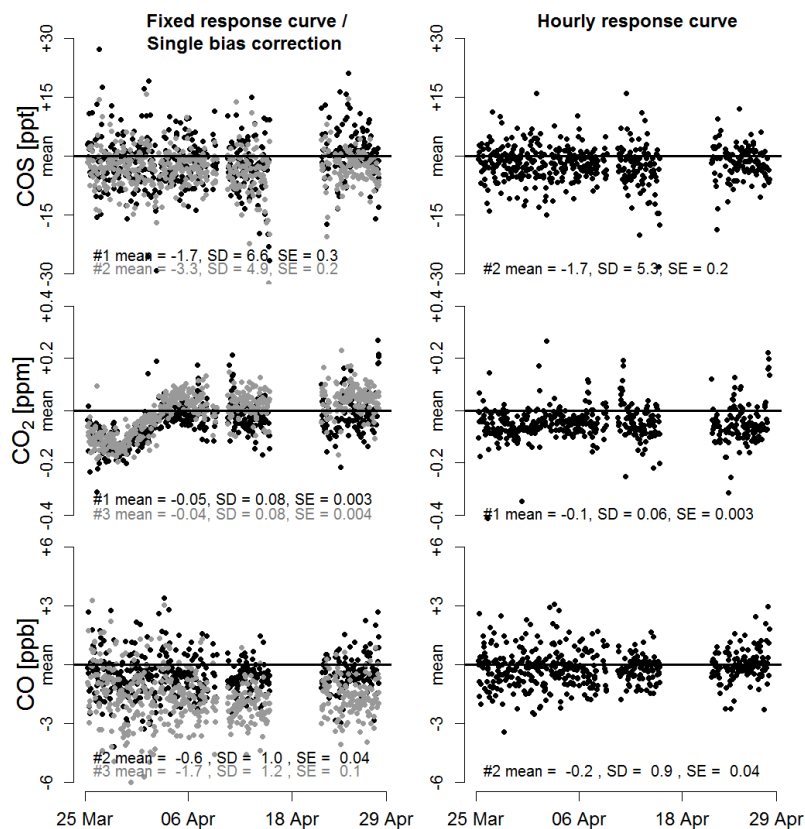


Figure 8 Mole fraction offsets in target standards after application of (left) corrections with a fixed response curve for CO₂ and CO that was determined in the laboratory (see Sect. 2.2.1), and with a single bias correction for COS, and (right) corrections with changing response curves determined from the hourly measurements of working standards. For the fixed response curve measurements of two target standards are shown, for the hourly response curve only one is shown, as two of the three working standards were needed to determine the response curve. **The mean deviation from the assigned cylinder value, standard deviation (SD) and standard error (SE) are given for each target standard. The cylinder numbers (#1, #2, #3) correspond with the cylinder numbers and assigned cylinders values given in Fig. 6.**

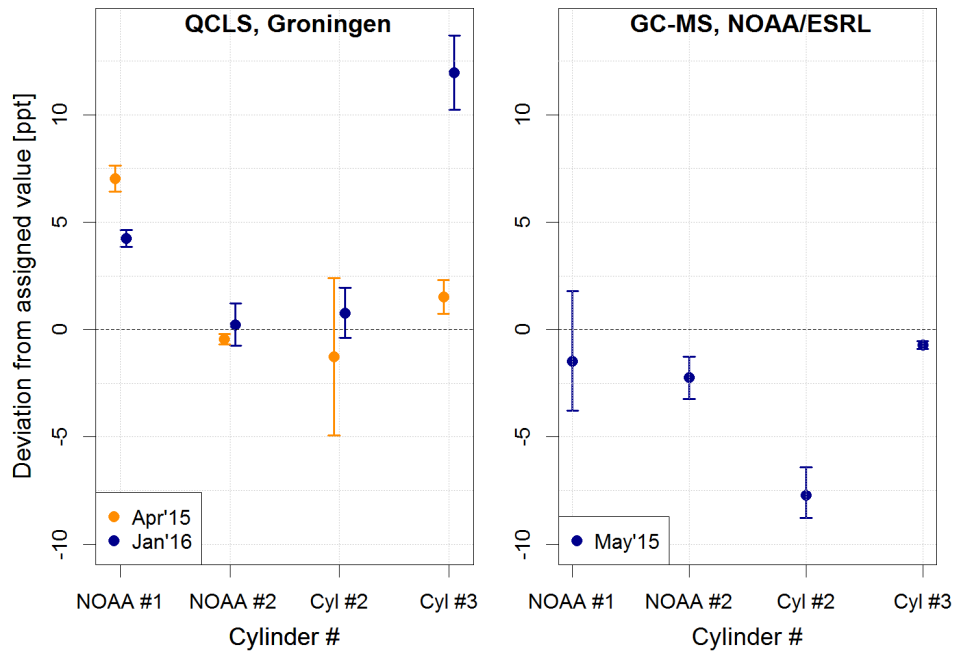


Figure 9 Flask sample measurements at the QCLS in Groningen (left) and GC-MS by NOAA/ESRL (right). Four paired flasks were filled with dry air from the two NOAA/ESRL calibration standards and two working standards, which were calibrated as in Sect. 2.2.2. The flask pair measurements are averaged and shown as the deviation from the assigned cylinder value. The first two cylinders are those calibrated at NOAA/ESRL and have an assigned value of 447.8 ppt (NOAA #1) and 486.6 ppt (NOAA #2). The last two cylinders were used as working standards in Lutjewad and have assigned values of 467.6 ppt (#3) and 455.5 ppt (#4). The error bars show the precision. For the QCLS two measurements were done: in April 2015 (orange) and in January 2016 (blue). The GC-MS measurement at NOAA/ESRL was performed in May 2015.

10

15

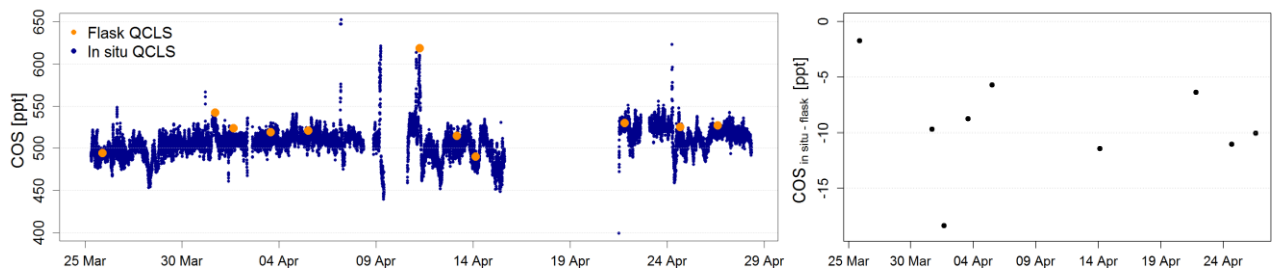


Figure 10 Minute averaged measurements of COS at the Lutjewad measurement tower (60 m) between March 25 and April 29, 2015 as measured *in situ* by the QCLS (blue), compared with dry air flasks samples measured at the QCLS (orange). The left plots show the time series, the right plots show the difference between the *in situ* and flask measurement.

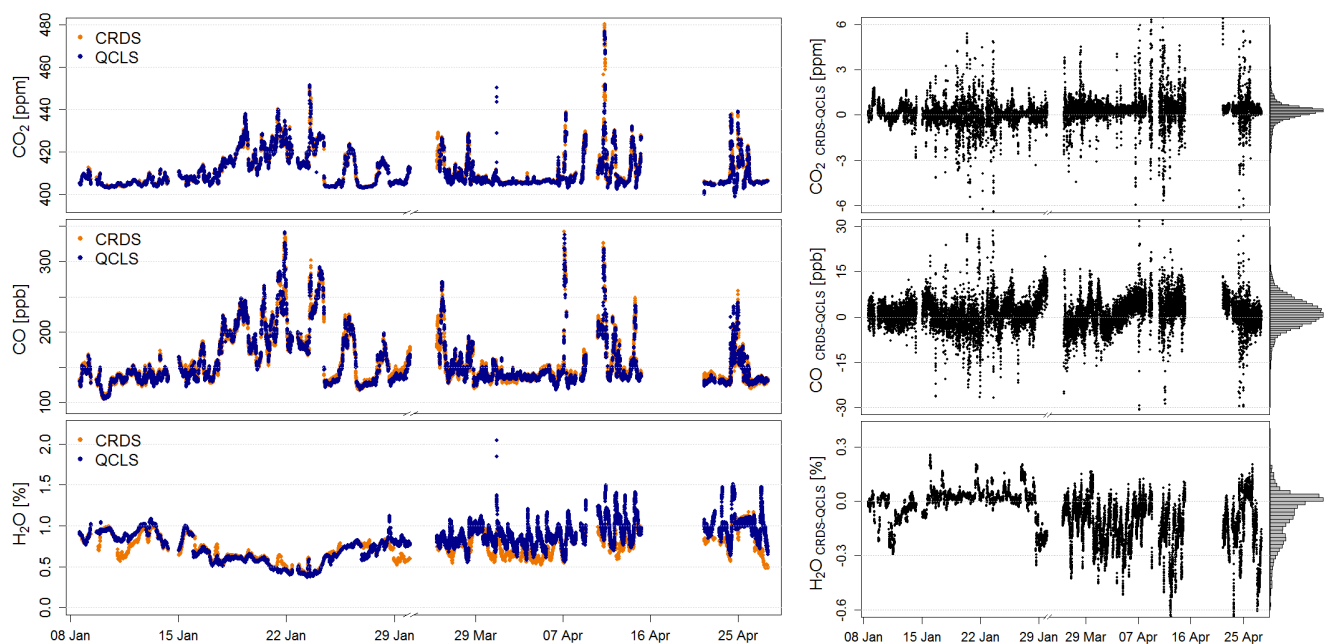


Figure 11 Minute averaged measurements CO₂, CO and H₂O of the Lutjewad measurement (60 m) tower between January 7 and April 29, 2015 as measured by the QCLS (blue), compared with measurements from a CRDS (orange). The left plots show the time series, the right plots show the difference between CRDS and QCLS measurements together with the frequency distribution of the difference, which is plotted as a histogram.

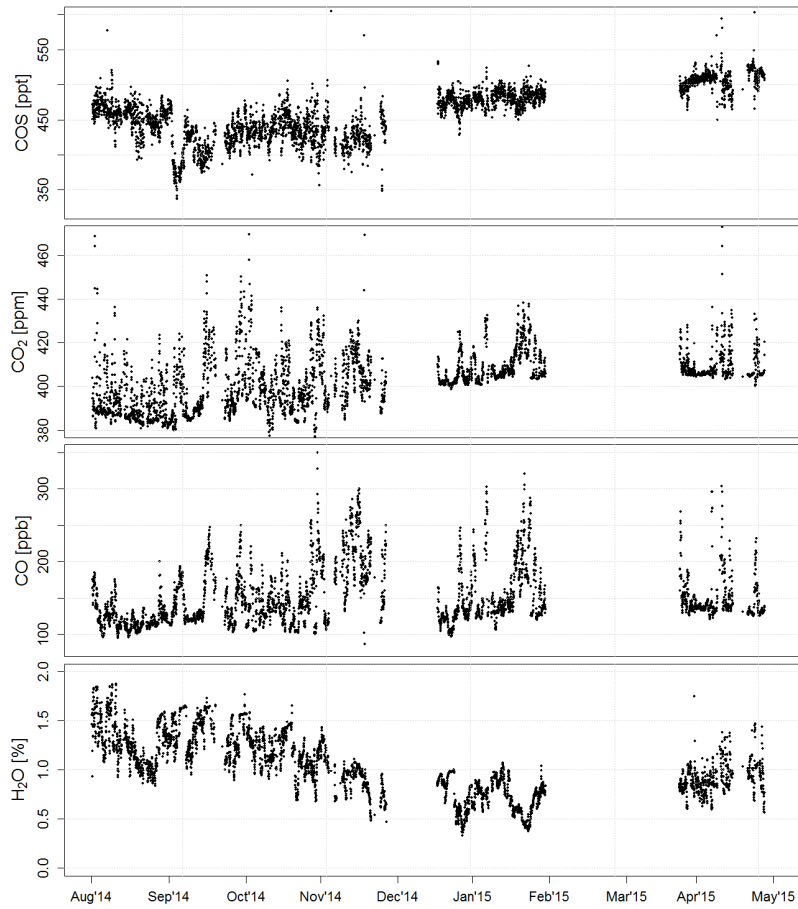


Figure 12 Hourly averaged measurements of COS, CO₂, CO and H₂O in ambient air at the 60 m level of the Lutjewad measurement tower as measured by the QCLS. Data are shown with corrections as described in the text.

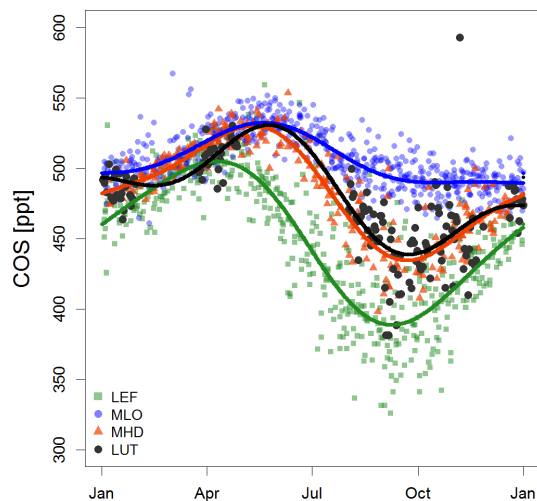


Figure 13 COS seasonal cycle of 4 sites: Wisconsin, United States (LEF), Mauna Loa, United States (MLO), Mace Head, Ireland (MHD) and Lutjewad, the Netherlands (LUT) where the data of the latter site is presented in this study. COS mole fractions for the LEF, MLO and MHD sites were measured from flask samples at a GC-MS by NOAA/ESRL (Montzka et al., 2007). The NOAA/ESRL data are shown as flask pair means from individual sampling events. All NOAA measurements are plotted as function of time of the year and cover a period between 2000 to 2015 for LEF, MLO and MHD. *In situ* COS measurements with the QCLS at the Lutjewad site during 2014-2015 are shown as daily averages (black). A two-harmonic seasonal cycle is fit through the data.