Instrument inter-comparison of glyoxal, methyl glyoxal and NO_2 under simulated atmospheric conditions

R. Thalman, M. T. Baeza-Romero, S. M. Ball, E. Borrás, M. J. S. Daniels, I. C. A. Goodall, S. B. Henry, T. Karl, F. N. Keutsch, S. Kim, J. Mak, P. S. Monks, A. Muñoz, J. Orlando, S. Peppe, A. R. Rickard, M. Ródenas, P. Sánchez, R. Seco, L. Su, G. Tyndall, M. Vázquez, T. Vera, E. Waxman, and R. Volkamer

Response to Referees

Anonymous Reviewer # 3

The authors present here a unique dataset of two intercomparison measurements at atmospheric simulation chambers, one at EUPHORE and the second one at NCAR. The measurements of glyoxal, methyl glyoxal and NO_2 with a large range of instruments are the largest intercomparison for these instruments so far. It is an important research topic for the scientific community and thus this manuscript is a very valuable result for the scientific community. Indeed such manuscripts are typically long and difficult to read, as they contain a lot of information and aspects. But the authors made a good structure and separation in subsections. The manuscripts fits perfect to the scope of AMT. I have several aspects which I think are not addressed sufficiently or not absolute correct. I recommend publication after correction and improvement of the manuscript.

Major points: In the manuscript extensive comparison of the measurement results is performed. But I agree with the Comment from Thomas Hanisco, that the investigation of measurement accuracy at atmospheric levels (e.g. glyoxal at 50ppt) is coming too short. In the conclusion even the impression arises that accuracy for glyoxal is sufficient good. These concentration ranges should be addressed in more detail. Also a more critical statement on the problems, especially in the conclusion, would be very useful. This would not degrade the work from the authors. In the UV/ VIS spectral range where glyoxal and methyl glyoxal are evaluated, overlapping absorption of water is present. The available spectra data from HITRAN have large uncertainties in this spectral range and large corrections were applied to the HITRAN H_2O absorption cross section in this spectral range for the different HITRAN data products. It is unclear to me why this major issue for the glyoxal and methyl glyoxal spectral data evaluation is not investigated in much more detail during this study. I understand that this is now not possible, but it should be stated in the Abstract, Section 3.2.4 and the conclusion, that during this study the effect of H_2O on the UV / VIS absorption measurement of glyoxal and methyl glyoxal was not investigated systematically and in more detail and thus no clear conclusion on the influence of H_2O absorption can be made.

We thank the reviewer for this thorough assessment of the paper, and the valuable comments. We have added a new Section 4.6 'Comparison of atmospheric glyoxal concentrations' that contains a new Table 5 with correlations of ambient glyoxal concentration (mostly below 300 pptv, and in all cases below 500 pptv), and for low concentration data from experiments in the absence and presence of NO_x and under dry and moist conditions.

The conclusion section now contains a summary statement on the outstanding areas for future study. We agree that it is worthwhile to investigate the diverse possible effects of H_2O in more detail. However, we respectfully disagree that 'no clear conclusion on the influence of H_2O absorption can be made', since the absolute humidity H_2O effect was investigated systematically. We have added a new Section 4.5 'Interferences from H_2O' to make our reasoning transparent for the choice of experimental conditions. A setup like EUPHORE does not provide good control of temperature and RH, which are initially coupled. Suggestions for future studies are now also mentioned in the Conclusion section.

More specific points:

p. 8583 l.28: "For glyoxal and methyl glyoxal the slopes varied by less than 12% and 17% (both 3-sigma) between inherently calibrated instruments (i.e., calibration from knowledge of the absorption cross-section)." I do not understand what you mean with 3-sigma for a slope. The slope is the slope of the linear correlation fit. What does than the 3-sigma mean?

3-sigma refers to the uncertainty in the slope as given by the weighted fits.

p. 8587 l. 17: "R=0.999972" is it the measured reflectivity? Is it the peak of reflectivity? To which peak absorption path does it relate in air/ vacuum under your measurement conditions?

We have added "measured" to the text to clarify that this is a measurement of the mirror reflectivity. This R corresponds to an absorption path of 18.4 km in air, and 32.5 km in vacuum.

p. 8588 l. 6: "empty cavity": what do you mean with empty? Vacuum?

"Empty" refers to absence of absorbers other than O_4 (clean air). The description of the equation has been expanded to read: "… in the cavity and compared to the retrieved O_4 signal at 477 nm." to reflect that the equation describes all absorbers.

p. 8588 Eq. (1): The absorption path L(λ) varies with wavelength lamda, also absorption cross sections vary with wavelength (this should be added). But the right side of the equation using the O4 SCD does not depend on wavelength. So this equation is in that way not correct. It is just an approximation for the wavelength where you evaluate the O4 absorption.

Good point. This equation has been modified to clarify the general nature of the wavelength dependence; and that the same equation can be used to demonstrate cavity control if evaluated at 477 nm (the center of the O_4 band). The equation has been split into 1a and 1b, 1a being the general expression and 1b the expression solving L(477 nm) compared to the fitted O_4 band.

p. 8588 l. 21: "...and perform DOAS fitting of multiple reference spectra simultaneously. Literature absorption cross-sections for glyoxal (Volkamer et al., 2005b), methyl glyoxal (Meller et al., 1991), NO2 (Vandaele et al., 2002), and O_4 (Hermans et al.,1999; Hermans, 2010) were used in fitting the spectra." – As it is written in other sections, I understand that the same literature absorption cross sections are used for the different instruments. Is this correct? But this is the only definition of reference spectra. It is not obvious from this section that the same references

from here are used for the different instruments. Also H_2O is not included here for the experiment with higher RH. I suggest making clear which instrument uses which reference spectra (e.g. in Table 1).

A common list of reference spectra was agreed upon. We have added an explanation of the common literature cross-sections in the initial description of instruments in Section 2.1 as follows: "All visible absorption instruments used the same literature cross sections for the retrieval of glyoxal (Volkamer et al., 2005b), methyl glyoxal (Meller et al., 1991), NO_2 (Vandaele et al., 2002), O_4 (Hermans et al. , 1999; Hermans, 2010) and water vapour (Rothman et al., 2009). A further discussion of the infrared cross-sections used by the instruments at the two different facilities is discussed in their respective descriptions and in Section 4.1."

p. 8588 l. 21: "The DOAS output in units of slant column density (SCD=concentration x c x L) was then divided by the path length to get concentration." – But the absorption path length depends on the wavelength. I can somehow follow how the analysis was done, but the way it is written, it is not clear.

The SCD is the output of the DOAS fit. The wavelength dependence is accounted for by modifying cross-sections with respect to a reference wavelength (see Thalman and Volkamer, 2010 for discussion of cross-section scaling factors and the retrieval, as well as the discussion of this process in the preceding sentences). The revised manuscript has been clarified that the cross-sections are scaled by the changing path length and provided as input to the DOAS fit.

p. 8588 l. 21: "Equation (1) was solved iteratively to account for self-limitation until the concentrations converge (either for NO_2 (experiments 3, 4, 7, 9 and 10) or glyoxal (exp 1 and 8))." – Why it should converge differently for NO_2 and glyoxal? The physical principle is the same.

Yes, the physical principle is the same for both NO_2 and glyoxal. But the conditions during experiments vary, and self-limitation is not always exhibited by the same compounds. NO_2 is self-limiting during experiments E3, 4, 7, and 9/10, and glyoxal is self-limiting during experiments E1 and 8. The language was modified in the revised manuscript to make this clearer.

p. 8589 l. 16: "LED peaking around 455nm" – Is the LED intensity stabilized (e.g. with temperature stabilization). The BB-CEAS data analysis algorithm relies on a stable intensity of the light source. Any intensity variation directly scales the observable concentrations and thus significantly increases the measurement error. This cannot be observed in measurements without absorber (zero drift experiments like shown in Fig 9 and 10).

The BBCEAS instrument's LED is temperature-stabilized on a Peltier cooled mount. However, the authors respectfully disagree that LED intensity variations necessarily cause a linear scaling in the retrieved absorber concentrations. Our experience is that intensity drifts cause spurious broadband offsets in BBCEAS spectra (akin to aerosol extinction). Retrieval of highly structured absorbers like glyoxal and NO_2 rely on the differential structure lying on top of any spurious broadband continuum, are surprisingly robust in the presence of intensity drifts (in our case, we can use a relatively high 6^{th} order polynomial in the BBCEAS fitting to filter out the broadband contributions).

p. 8589 l. 18: "peak reflectivity= 0.999817 at 462 nm" – measured? To which peak absorption path does it relate in air/ vacuum under your measurement condition.

The text has been update to note that this is the measured mirror reflectivity. As was described for the CE-DOAS instrument this is the characteristic of the mirror coating (see Section 4.4 for the estimate path length of the cavity under clean conditions in air which is approx. 5km using equation 1).

p. 8590 l. 9: "typically 6th order" – This is very high for a spectral analysis and cannot be explained by the aerosol absorption. Only the missing ozone absorption could explain it (as stated later). Also other instrument issues could cause such problems. Thus clarify this already here.

We have tested various orders on other (but comparable) datasets: 4^{th}, 5^{th} and 6^{th} order polynomials return the best spectral fitting statistics, with 6^{th} order being marginally the best of these three. It's true that using a high order polynomial does help to filter out broadly structured absorbers like O_3 and biacetyl. We have found a 6^{th} order polynomial also helps to mitigate the effects of small drifts in the LED's output which often add some curvature at the long and short wavelength extremes of BBCEAS spectra.

As noted elsewhere in this response; a wide analysis window is necessary to make use of a high degree of polynomial, and this approach is preferred for BB-CEAS retrievals. For CE-DOAS retrievals (which are narrower) a lower degree of polynomial is used. See also our response to Chan Miller.

p. 8590 l. 14: "instrument's spectral resolution (between 0.09 and 0.13nm half width at half maximum)". This high spectral resolution with this spectrometer indicate an under sampling of the instrument function (min. 5 to 10 channels). Even if a used asymmetric line shape function is derived from 20 lines this is problematic, as the line shape function is changing significantly over the spectral range of the applied spectrometer. Thus not necessary the "real" line shape function for the fitting window is derived. Is this error source evaluated?

Indeed, there are 5 to 10 channels across the full width at half maximum of our instrument function; this sampling rate is typical also of DOAS instruments. We use an empirical instrument function based on three parameters: center wavelength, width, and an asymmetry parameter used to skew the lineshape (details to follow in and instrument paper in preparation "Daniels et al"). We fit instrument functions to approx. 20 lines from emission lamps recorded across the spectrometer's bandwidth. We then plot the results for the width and asymmetry parameter versus wavelength. We typically find that these parameters vary smoothly with wavelength; hence we fit plots of the fitted line parameters vs wavelength with cubic polynomials to infer width and asymmetry parameters at all wavelengths across the spectrometer's bandwidth.

We have fitted representative BBCEAS spectra using both the asymmetric lineshapes (above) and more conventional Gaussian line shapes. The retrieved concentrations of highly structured absorbers (glyoxal and NO_2) are the same within their statistical fitting errors. However the asymmetric lineshapes do a

better job of removing the spectral structure due to glyoxal and NO_2, and hence yield smaller residuals and smaller fitting errors for any weak absorbers that lie underneath the glyoxal and NO_2 spectral structure.

p. 8590 l. 14: "The instrument was subject to small long-term drifts over the _12 h duration of the Allan tests that degraded the achievable precision." Alan test at zero concentration does not show errors due to intensity drift of the light source, as this will be in BB-CEAS a scaling of the derived concentration. But if concentration is zero, no scaling effect will be visible.

See our response to previous comment on p. 8589 line 16 about LED drifts. These (generally) do not have a linear effect on the retrieved absorber concentrations – although LED drifts can produce spurious (generally broadband) features in BBCEAS spectra that retrieval algorithms can try to fit as an absorber's signal.

We agree that it would be desirable to conduct Allan analyses using a stable supply of an absorber diluted to a stable concentration. However, there are substantial technical challenges involved in producing a very stable test gas mixture; in practice such an Allan test is likely to reveal instabilities/drifts in the source, rather than in the measuring instrument. The aim of conducting Allan tests at zero concentration is to examine instrument stability in isolation. This test necessary produces the lowest (i.e. most optimistic) sensitivity values, corresponding to quantifying an absorber against a zero background. This approach has been widely used for cavity instruments (e.g. J M Langridge et al, Rev Sci Instrum, 79, 123110, 2008; N.L. Wagner et al, Atmos. Meas. Tech, 4, 1227, 2011; T. Wu et al., Appl Phys B, 106, 501, 2012; W. Zhao et al, Analytical Chem, 85, 2260, 2013.)

p. 8591 l. 15: "The overall accuracy of the BBCEAS concentration measurements is estimated to be 7% for glyoxal and NO2 and 10% for methyl glyoxal." – The noise in the spectrum will be the dominant error source at very low concentrations. It defines a minimum error for the measured components. Thus an error in percent is not absolute correct as it would give a much better accuracy for very low concentrations, than the correct value.

The above percentage values relate to the accuracy, i.e. to systematic sources of errors on the BBCEAS measurements. It is certainly true that statistical (i.e. spectral fitting) errors, not these systematic errors, dominate the total error at low absorber concentrations. For this reason, the error bars shown on the BBCEAS measurements throughout this paper are the combined systematic and statistical error. See also our comments to Chan Miller about the BBCEAS error bars in Figs 4 & 6 (Now Fig. S3). A full discussion of the fitting errors for the BBCEAS instrument will be given in our subsequent instrument paper (Daniels et al).

p. 8592 l. 11: "using dry nitrogen as a carrier" – Is it comparable to the measurements?

Yes. There is no phosphorescence from the carrier gases.

p. 8592 l. 18: "tuning on the scale of the vibro-rotational absorption spectral features of glyoxal (~0.06 nm)" - At which wavelength? What is the phosphorescence wavelength? Please clarify this at the beginning of the instrument description. This would also be useful to understand the

different filters on p. 8593. It is often not clear if you mean the laser light or the phosphorescence light.

The excitation wavelength is 440.014 nm and the detection window is 520 (±5) nm as specified on the next page in the section the excitation wavelength has been moved to also be mentioned at the start of the section and the detector description has been clarified to refer to the phosphorescence photons.

p. 8593 l. 1: "The variability of the alignment is reflected in variability of the calibration factors." – Please clarify. The misalignment should not change the absorption path in a white cell. So is this just an intensity variability which would require a new calibration?

A change in alignment has two effects:

1. Throughput, i.e., how much light is scattered from mirrors or absorbed/scattered by baffles as opposed to exciting glyoxal.
2. the position of the beams relative to the focal point of the PMT assembly

In effect the amount of light for excitation and the collection efficiency are affected by alignment changes. Normalizing by laser power helps address this, but cannot do so perfectly.

p. 8602 l. 16: "It should be noted that the W-DOAS instrument is affected by the distortion of the light beam during the flushing of the chamber" – How should a distortion of the light beam affect the correlation? The distortion should reduce light intensity but not the absorption path and thus only slightly increase the measurement error, but not the value itself.

The injection of large volumes of gas to flush the 200m3 volume reactor is into the optical path of both instruments, and introduces turbulence that can lead to distortions of the light beam worsening the detection limit, and also making an inhomogeneous air mass, which varies the concentration of the compounds measured.

p. 8603 l. 14: "Deviations in the SPME concentrations were large but appear to be unconnected to the high NO_2 levels in the chamber. For both CE-DOAS and BBCEAS (Fig. 4) we do not find significant bias, i.e., an upper limit change in glyoxal due to NO_2 is derived as ±200 pptv glyoxal in the presence of 200 ppbv NO_2 (or 1 pptv glyoxal/1 ppbv NO_2)." – I cannot follow this conclusion. I think the plot S5 show a clear connection of glyoxal with NO_2. I think this is an important analysis and should be part of the main paper. A bias of 200ppt for glyoxal would be a significant value for atmospheric concentrations.

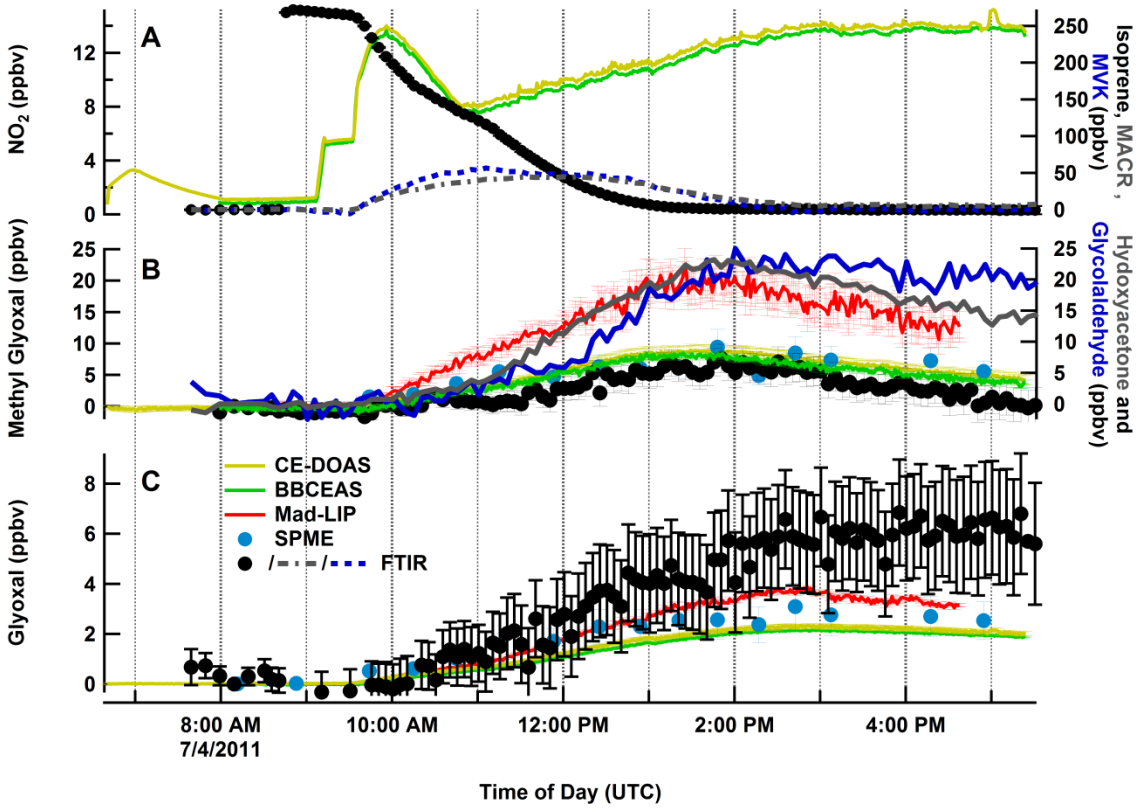
An NO_2 concentration of 200 ppbv is not a common occurrence in the atmosphere. Section 4.6 'Comparison of atmospheric glyoxal concentrations' discusses limitations at low glyoxal concentrations and under atmospherically meaningful conditions, including for NO_x. We have added Fig. S5 and S6 to illustrate the effect of NO_2 in ambient air.

p. 8604 l. 12: "before and after HONO addition" – The structure of this section is very confusing. Why the addition of HONO is mentioned not at the beginning when the experiment is explained?

The section has been reworded to mention HONO earlier and simplify some of the introductory language.

p. 8604 l. 28: "However, the variations in slopes were somewhat higher, i.e., 0.94–1.5 (glyoxal), and 0.7–2.2 (methyl glyoxal), while most instruments agreed within 30% for both species (see Table 3)." - If variation is so high, could you provide a plot in the manuscript?

The following new Figure was added as Figure S4 to the Supplement.



p. 8605 l. 26: "The RH varied between 45% and 58 %" - First of all this variation is very small to investigate an interference. Second, for the spectral interference not RH but concentration is important. Third, it is not clear to me how from the performed measurements an influence of water on the measurement of glyoxal and methyl glyoxal should be observed. And this should actually be done for concentrations which are typical in the atmosphere. I cannot see how an influence in the range of 100ppt for glyoxal can be excluded from the presented measurements just by the intercomparison of the different instruments at typically much higher concentrations. So how the authors derived the influence of water absorption? Why is there no W-DOAS data for this experiment?

We agree that absolute humidity is the important parameter, and have added a new Section 4.5 'Interference from H_2O' and a new Table 5 for low concentration glyoxal correlations. The range of

absolute humidity for the comprehensive dataset was actually substantially larger; the reviewer only refers to E6. This is now made transparent. See also response to reviewer Chan Miller.

There was no W-DOAS data for this experiment as the mirrors were covered to protect them from deposition of aerosols on the dielectric coatings and there was no time to change out the mirrors for the aluminum coated set and bring them back on line for the last several experiments.

p. 8606 l. 1: "The slopes of correlations (Table 3) varied between 0.95–1.5 (glyoxal), 0.68–1.83 (methyl glyoxal),..." - This strong variability may indicate serious issues. It is linked to the question above.

The numbers quoted by the reviewer need to be regarded within the error bars of respective techniques. The two instruments/methods that give the largest deviations are glyoxal- SPME (which has limited signal to noise at the low concentrations investigated here), and methyl glyoxal- Mad-LIP (which has a known issue with conflicting signals from glyoxal on the methyl glyoxal measurement)).

The issues actually are much smaller if only data is used that has the precision to assess water effects. We have added a new Section 4.5 for discussion of issues related to water. We have also added a new Section 4.6 'Comparison of atmospheric glyoxal concentrations' that contains a new Table 5 with correlations of ambient glyoxal concentrations, and for a subset of low concentration data from all experiments in the absence and presence of NO_x and H_2O.

Mentioning of the issues and ways to further investigate them has also been included in the conclusion section.

p. 8606 l. 16: "Interference from O3" - This title seems to be confusing as not measurement interference directly due to ozone is investigated but rather the glyoxal production due to ozone. A second point to this chapter: You investigate mainly glyoxal production from ozone in Teflon tubing. But one of the mayor issue people worry is glyoxal loss in lines, walls filters etc. especially if they get wet. Why this is not further investigated? The authors themselves used heated sampling lines before to avoid such losses.

Interference does not necessarily mean 'spectral' interference. We refer to the broader definition of interference as 'something that obstructs or hinders'. Ozone producing glyoxal in the lines is an example for such an interference.

The first sentence of the paragraph has been changed to read:
"Experiment E5 tested the interference of O3 directly via spectral interference with absorption measurements and/or indirectly by either production of glyoxal on reaction with Teflon (walls of the chamber or sampling lines) or other VOC's in the chamber."

The choice of parameters that were varied in this study includes H2O effects, which were the primary reason for why we have used heated lines before. The new Sections 4.5 and 4.6 make our rationale for the choice of experimental conditions in this study transparent.

p. 8606 l. 16: "CE-DOAS and BBCEAS instruments changed the lengths of their sampling lines to attempt to observe any change in the measured concentration." – Are these all new clean teflon lines? Can this be adapted to teflon lines used e.g. for few days at ambient conditions?

Yes, the BBCEAS and CE-DOAS instruments did this test with a brand-new Teflon lines. The suggestion of the referee of ambient exposure of the lines is sensible, but at this stage, we can't re-do any experiments using lines that have sampled ambient air. We explained that they are new lines in the text.

p. 8607 l. 17: "Fig S2" – This seems to be a rather important data set and should be part of the main manuscript. Fig 9 and Fig 10 could be moved to the supplement materials.

The results from Fig. S2 (now S8) underlie Table 4 and are reflected in the main text. Fig 9 and 10 are important to the overall discussion of precision and accuracy (Sect. 4.2). It is impossible to have all aspects of interesting data in the main text for a study like this. We feel that this particular Figure is better placed in the Supplement text.

p. 8609 l. 6: "Direct comparison of the EUPHORE and NCAR IR spectra showed a factor of 0.78 difference, which was traced to a near identical correction factor that had previously been applied to the EUPHORE-IR spectrum (see Sect. 2.1.8)." – I could not find in the section 2.1.8 any explanation about this scaling factor. It is also not clear to me how this scaling can be explained. Please explain or give according references.

We have removed the reference to Sect. 2.1.8 here. The origin of the factor of 0.78 is of rather marginal interest to the conclusions of this study. It was necessary to understand this factor in order to discard it upon further inspection - this is what the text states.

p. 8609 l. 21: "in slopes between PTR-ToF-MS and CE-DOAS of 0.95 ±0.03; this is essentially unity at the 95% confidence level." – The subordinate clause >this is essentially unity at the 95% confidence level< is obvious if you define your confidence level in such a way. Thus it is confusing and should be removed.

This has been removed and replaced with the addendum +/-0.03 (~1 at 95% C.I).

p. 8610 l. 5: "LOD=3 x 1-Sigma variability+background" – what do you mean with background? It is unclear what value you are describing.

This notation is the basic definition for limit of detection as summarized at textbook level (see Harris, Quantitative Chemical Analysis 10th ed., p. 103 for example). The revised manuscript clarifies that 'background' refers to the measured value of a blank sample. The associated uncertainty is assessed through Fig. 9 and 10, and is now reflected in Table 4.

p. 8610 l. 15 "Any deviation from pure white-noise residuals can be accessed from multi-channel sensors, and provides additional information to assess LOD from a perspective of "accuracy"." – Also the size of the white noise give additional information to access the LOD.

Yes, this is true – however, here we are discussing 'deviations from pure white-noise'.

We have added a statement near the end of Sect. 4.2 that reads: "The size of white-noise residuals observed by CE-DOAS can be understood in terms of the measured photon fluxes, and provides addition information to assess LOD and accuracy (Fig. S9, Supplement text)."

p. 8610 l. 20 "We used Eq. (7) to calculate experimental LODs using the 1-sigma variability of data from the overnight dilution experiment on 5–6 July 2011 (E8b; see Fig. S2 in the Supplement)." – But this method ignores systematic offsets of an instrument what clearly increases the LOD. Thus the calculated LODs are too optimistic, as offsets can clearly be observed already without interfering gases in Fig. 9 and 10. At least these offsets should be included and influences from other interfering gases on LOD should clearly be pointed out.

This is a good point. Table 4 was revised to include these offsets, and Section 4.2 now includes some discussion about the relative contributions of offsets and variability contributions to LOD. In addition, a graphing error in the CE-DOAS panel of Fig. 9 was corrected to show the mean of the distribution (-1.5 pptv) falls within 1 σ (3.5 pptv) of 0. Additional sensitivity tests were performed to assess instrument drift for CE-DOAS that employed different reference spectra (variable time difference) to analyze the period of interest here, and a sentence was added in Section 4.6. The influence of other interfering gases on LOD cannot be quantified in any one way that was representative for all conditions. But the new section 4.6 discusses limitations at low concentrations. At high concentrations we believe our assessment of accuracy in Table 4 is appropriate as is.

p. 8610 l. 23 "listed together with LOD values submitted with their measurement data by the operators of the various instruments." – Are the operators derived LOD are calculated in the same way?

The operators derived their values from their own data according to their various methods outlined in the literature and submitted them with the data to a central archive. This was then compared to the method outlined. The operator reported LOD does not include any offsets ('background' in Eq. 7), as determination of offsets needs knowledge that can only be obtained from comparison with other techniques.

p. 8611 l- 19: "from the overnight dilution experiment" – Provide also experiment number from Table 3.

Text has been changed to read: "from the Experiment E8b, the overnight dilution"

p. 8614 l- 15: You start here explaining a different experiment. However this is not clear in this chapter. I suggest rephrasing the sentence to make it clear to the reader.

This paragraph has been reworded to first identify the experiment being discussed.

p. 8614 l- 26: "bias due to NO2 as ca. 1 pptv glyoxal/ppbv NO2 (Fig. 4) and 5 pptv methyl glyoxal/ppbv NO2" – How these two values derived from the plot? Please be more specific.

These values are derived by dividing the maximum offset by the maximum observed NO_2 and is described in Section 3.2.1.

p. 8615 l- 7: "The combined effect is a decrease of a factor of 16 in sensitivity." – This seems not to be correct. The loss in sensitivity due to the loss in light is only sqrt(4) = 2, and the loss due to the reduction in light path is ~4. Thus the total loss in sensitivity is 2x4=8.

The error in the calculation has been corrected in the manuscript.

p. 8616 l- 1: "Alpha-dicarbonyl" – I suggest to use glyoxal and methyl glyoxal to be more systematic in the manuscript and not to confuse the reader who is often not familiar with the name of Alpha-dicarbonyl.

We have removed "Alpha-dicarbonyl" from the section heading.

p. 8616 l. 3: "detection of glyoxal at ambient mixing ratios in urban, semi-polluted, biogenic, arctic and marine environments. In most urban environments the glyoxal detection by in situ UV-vis absorption techniques is feasible,.." – Please be more specific. What are typ. concentrations observed in these areas. How do the concentrations in this study relate to these concentrations? From the given data it is not obvious that the instrumentation is feasible to measure these concentrations as your comparison is mainly at high concentrations. Also a detailed analysis of water interference was not performed.

We have added typical concentrations in this paragraph. Also, the new Sections 4.5 and 4.6 make transparent how our results are useful to inform ambient measurements, including the water interference. The section now reads as follows: "…detection of glyoxal at ambient mixing ratios in polluted urban (1.5 ppbv (Volkamer et al., 2005)), semi-polluted rural air (100-500 pptv, (Washenfelder et al., 2011; Knolte et al., 2014)), forests (0.25-1.5 ppbv (Huisman et al., 2011)), marine boundary layer (20-50 pptv (Sinreich et al., 2010; Coburn et al., 2014; Volkamer et al., 2015)) and the free tropospheric environments (3-60 pptv (Lee et al., 1998; Baidar et al., 2013; Volkamer et al., 2015))."

p. 8617 l.1: The conclusion should be critical. It should of course state what was done, what is working etc. But it should also point out where still significant problems are, what should be done to minimize problems and which investigations are still needed.

The statements sound like there are mainly no major issues for the measurements. This can lead to misinterpretations. For example are the results of cross sensitivity not critical enough. They are maybe true for high glyoxal and methyl glyoxal values but not for very low values like typically present in the atmosphere. Also several experiments do not allow to give a clear statement like the observation with RH. So in each point of the conclusion it should be clearly stated for which conditions the specific conclusion is valid.

A critical statement has been added near the end of the conclusion, to help guide future work. The language of individual conclusion bullets has been edited to ensure that conditions are transparent. The added Table 5 informs conclusions at low concentrations for glyoxal.

p. 8618 l. 21: "Future studies should further investigate in detail the effect of O_3 and H_2O at very low concentrations of Alpha-dicarbonyls (<20 pptv) and high relative humidity (>80% RH), when losses/formation of Alpha-dicarbonyls in sampling lines or to/from aerosol filters are likely to be more relevant." – How do you come up with the values of 20ppt and 80%RH? How can you conclude that this is not already a problem at higher concentrations?

We have revised this statement to be consistent with the data supported in the new Section 4.6.

Supplement l. 69: "...Mirror Reflectivity (±2%), ..." – this would implement that the mirror reflectivity is measured with this accuracy. But how accurate is that measurement and how it is performed?

The mirror reflectivity is measured to this accuracy. It is performed from the differential scattering of nitrogen and helium as outlined in the experimental section. Further discussion can be found in two supporting recent references that describe measurements of O_4 absorption cross-sections (Thalman and Volkamer, 2013), and a comparison of Rayleigh scattering cross-sections with refractive index theory (Thalman et al., 2014).

Supplement l. 81: "the fit error for the SCD_{gly} is on the order of 15%..."- But here you do not recalibrate the light path like described in section 2.1.3. The measurement from O_4 and the scaling would introduce an additional, not negligible, error source.

In the high NO_2 experiments path length scaling is not done with O_4 fitting as the large NO_2 signal completely obscures the fit of O_4. As was described in section 2.1.3, an extinction fit (CEAS method, see Equations in Fiedler et al. 2003 and Washenfelder et al. 2008) for NO_2 was used to then calculate the path length for the glyoxal. This and any iteration is still smaller in the relative error than the fit error of the glyoxal under these conditions, as is also transparent from the discussion of error sources in the last paragraph of the Supplement text.

Supplement l. 83: ".. the fit error is 1.5- 2.0% over the full range of glyoxal concentrations investigated in absence of interfering species" – This makes no sense as you should have a minimum error due to the noise in the spectrum.

This number refers to high signal to noise, and does not factor in the cross-section uncertainty (this is done later – see last paragraph). For a discussion of error sources at low signal to noise see Section 4.6.

Supplement l. 119: "The overall uncertainty in the CE-DOAS calibration is 3.5%,..." –This error budget seems to miss the recalibration error from O4 (see above).

No – this statement is as intended, and does not miss any recalibration error from O_4. The CE-DOAS performs online measurements of O_4 at ambient conditions (low NOx) with a precision of about 1%.

When we compare the calculated O_4 column density from the mirror curve to the measured O_4 the agreement is better than 1%. The Rayleigh scattering cross-sections have been confirmed within 0.2-0.4 % to agree with calculations from refractive index theory. As stated near the end of the Supplement text: "…

Table 3: Why experiment E9 is missing?

Experiments E9 and E10 are at stable concentrations of glyoxal or methyl glyoxal and therefore correlations are not computed as there is little to no variability in the measured concentration by design (and therefore the slope/intercept are not well constrained). Correlation values for this data will carry very little meaning as the purpose to investigate the spectral effects of NO_2.

Table 4: Accuracy in % is not absolute clear. Please give reference to the section where a description of the calculation is given.

The footnote (a) specifies the variation in the fitted slope as the accuracy. We have added a reference to the section (4.2) in the footnote.

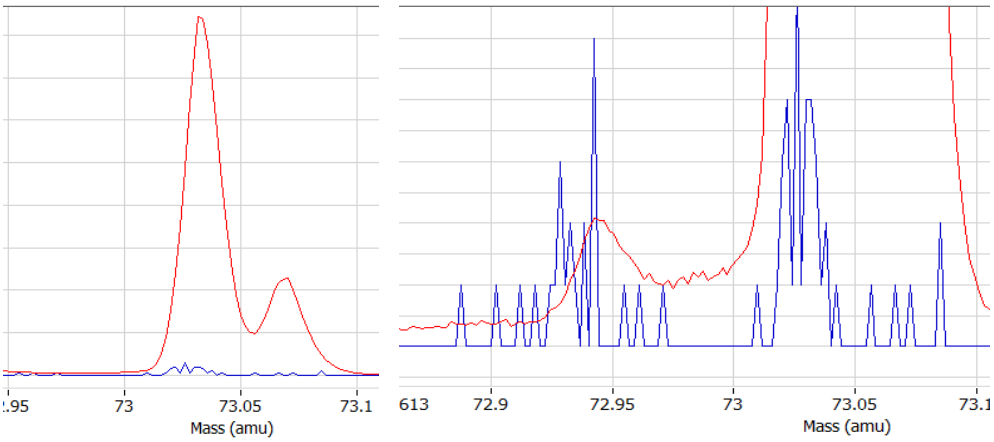
Fig. 1: Where is the FT-IR light path?
The FT-IR light path is located between the two sets of mirrors across the middle of the chamber and perpendicular to the W-DOAS light path (as shown in Fig. 1, and noted in the figure caption).

Fig. 2: Use same labels for the plots (small / capital letters). What do you mean with "Units of fit intercepts"? – The concentration? You use pressure and temperature of the camber to convert to pptv. But these values can be very different in the instruments outside the chamber. Was this considered? The PTR-MS sees a clear step at 20:35 which was not seen by the other instruments. How this could be explained?

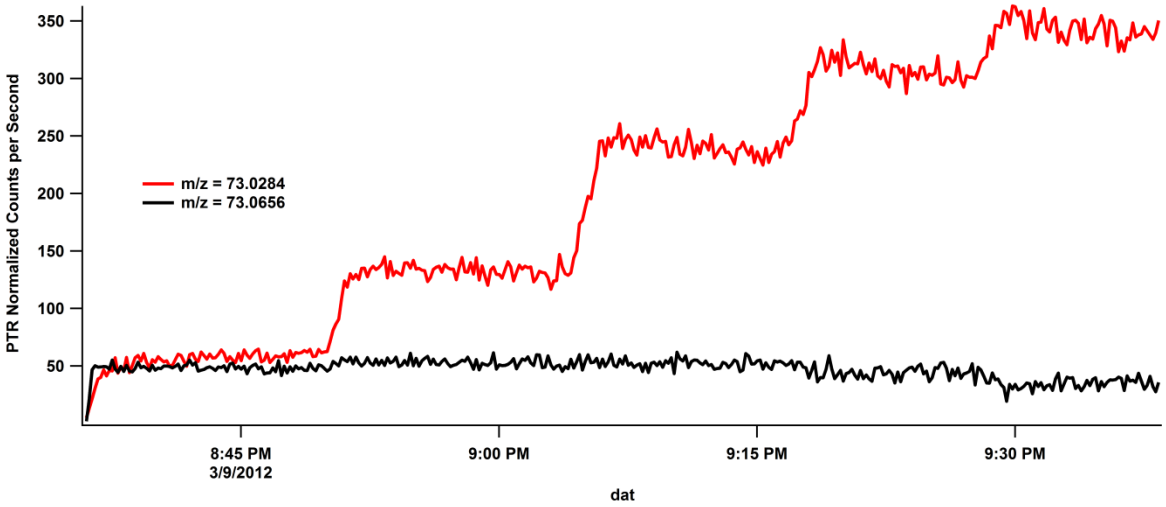
Yes – density effects are accounted in these calculations. The figure refers to the correlation data from the NCAR experiments. For these plots, to keep the consistency of the units of the EUPHORE experiments I have scaled the axes to mixing ratios. All correlation fits for NCAR data are done in concentration units, with external instruments scaled back to the measured temperature and pressure inside the chamber. While the slopes are relative (And do not need adjustment), I found it necessary to note why the units of the intercepts differ for this chamber compared to all of the others. The reason why the NCAR data won't be converted back to mixing ratio is that the number of experiments pooled is large and that the pressure throughout the experiments was changing. The FT-IR which was installed in the chamber itself served as the reference originally and measured absolute concentrations (relative to the fixed volume of the chamber) and the CE-DOAS and PTR were scaled back to the chamber volume for comparison.

At EUPHORE, the density differences between instruments housed inside the chamber, and outside the chamber, on and below the measurement platform have been accounted for. All instruments reported volume mixing ratios, which are independent of the measured temperature as long as each used the correct operating conditions to do the mixing ratio conversion. The W-DOAS and FT-IR used the real chamber temperature and pressure to calculate the air density used in reporting the mixing ratio inside the chamber.

The background in the PTR signal for methyl glyoxal comes from the interference of the neighboring mass peak as shown in the figures below, with the methyl glyoxal being the larger peak and the interfering peak to the right. The zoom on the baseline shows the real-time signal, where the instantaneous signal is integrated between a range of m/z.



The methyl glyoxal peak and the water cluster interference peak looked as follows:



If we subtract the interference (73.0656) from the methyl glyoxal then we get a new Figure S2 which gives a correlation of the corrected PTR data with a slope of 0.93(7) and an intercept of -3(10) x 10^{11} molecules cm^{-3} and an R^2 of 0.9996 relative to CE-DOAS.

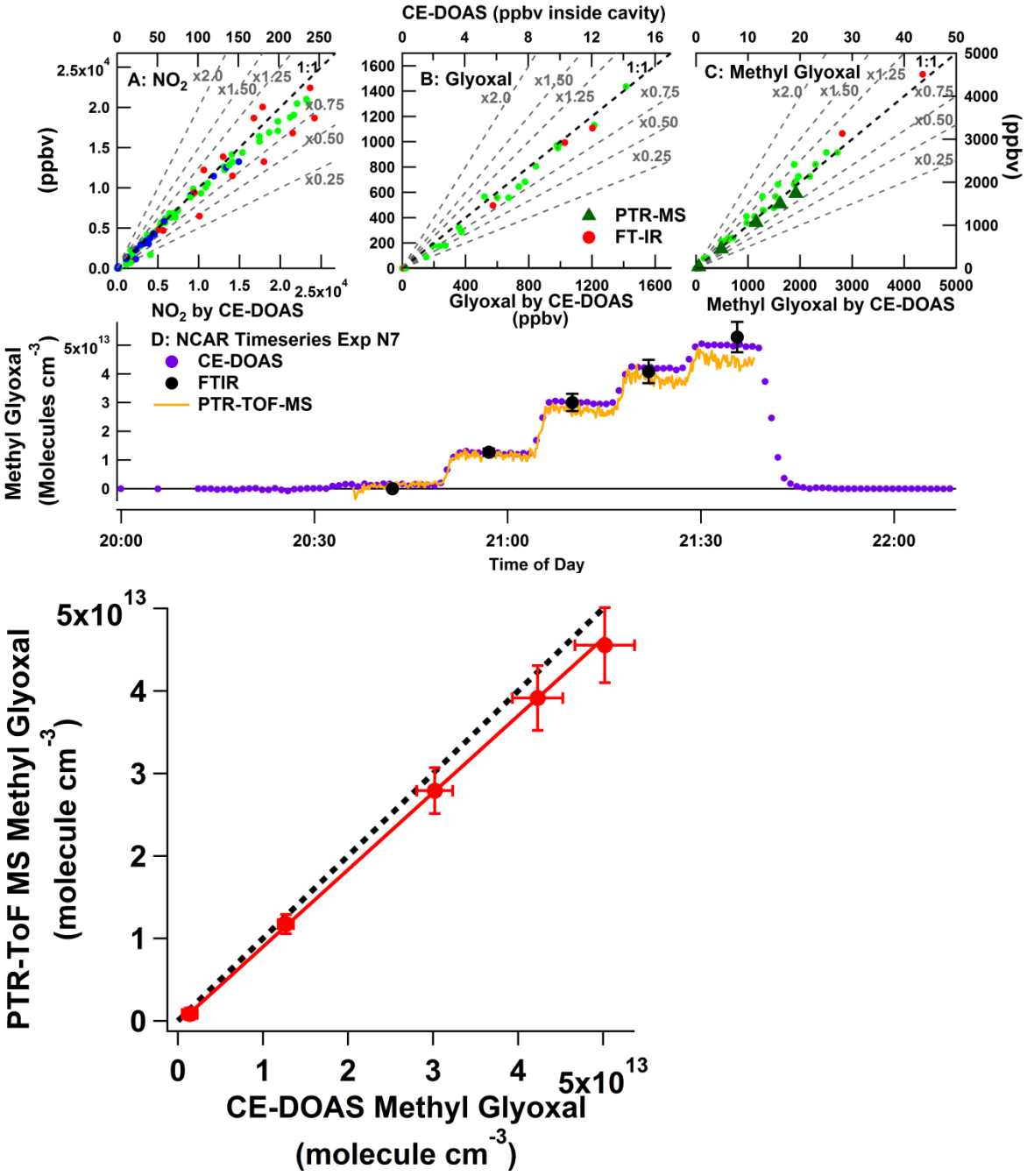


Fig. 3: The range below 300ppt is the most interesting, as typical ambient concentrations are below this value. However on this shown scale no information on this most relevant range is given. Even if several instruments have not sufficient accuracy, those who have could be plotted.

Discussion has been added as part of the new Section 4.6. Shown below are the values for low concentration glyoxal comparison on 5 July between BBCEAS and CE-DOAS with fit coefficients of (slope/intercept) 0.971(4)/-0.019(2).

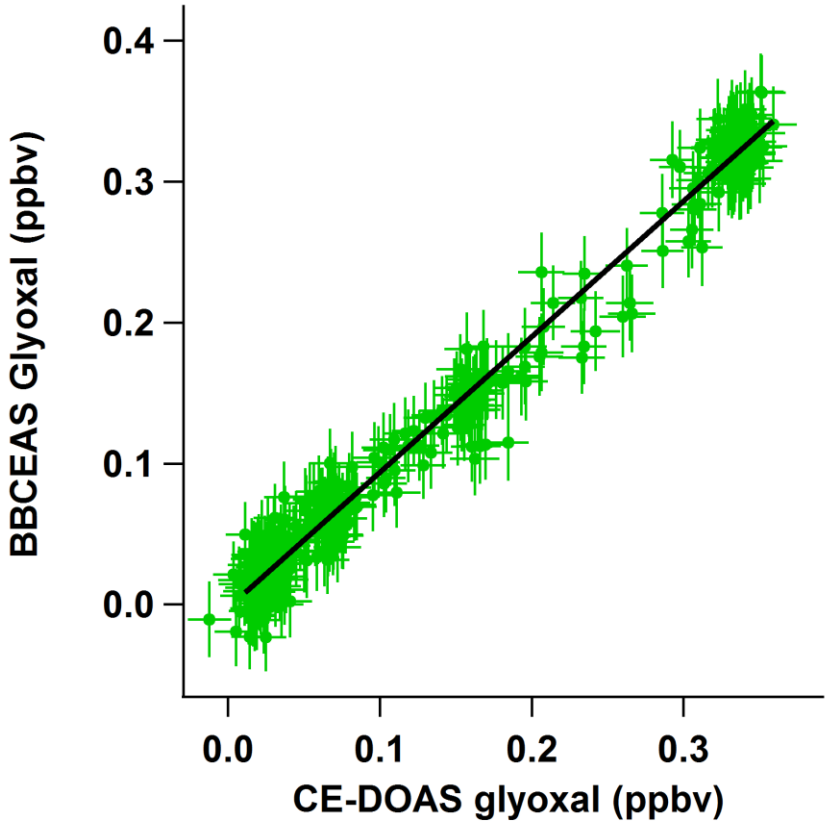


Fig. 4: Units for NO_2 are missing. A clear correlation to NO_2 is visible even if it is different for the different instruments. What is the implication for real atmospheric background levels of e.g. 50ppt?

Missing units have been added to the figure.

Fig. S5: In figure A: a clear dependency and trend for higher NO_2 levels can be found, thus the sentence in the caption seems to be wrong.

We agree that while there appears to be some trend, it is not apparent whether (1) it has any significance relative to the uncertainty or (2) the reason for the potential bias. We have removed the last sentence of the caption to match the conclusion in the main body of the paper.