Reply to Referees #1 and #2

We would like to thank both referees for the helpful comments and suggestions on our manuscript. For clarity, the referee comments and suggestions are italicized, followed by our replies.

Language style and typos as well as smaller issues were corrected and are not included in the following reply. The revised manuscript was proof read by a native English speaker.

Referee #1

A few times "mixing ratio" is written, but mostly "mole fraction". Use uniform terminology.

We now use mole fraction throughout the manuscript.

Most of the figures are somewhat too small. They need to be enlarged to make it easier for reader to see detail.

We enlarged the figures.

Page 4736, Introduction: References should be given for the older techniques.

We included references for the older techniques in the revised manuscript.

p. 4737, line 1: "a number of comparison studies …". Since only two ref. are provided, one should write "(e.g. Ou …)" or say "and references therein.

Done.

p. 4748, line 4: A possible source of error with the VURF instrument could be a difference between the zero as determined by the instrument during the zero/cal. cycle and the reading when CO-free (zero) air is analysed as a sample. Has this been tested?

This has not been tested during the period of the comparison; however, this would be a potential reason for the bias, although it would be more likely to have a negative bias (in case that not all CO is removed during the automatic zero calibration).

p. 4754: With reference to the "WMO/GAW Glossary of QA/QC-Related Terminology" (M. Steinbacher among the editors): Is it correct to use "precision" in quantitative expressions?

We think it is correct. The notes in the WMO/GAW Glossary of QA/QC-Related Terminology say that ‘Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement’.

p. 4755, caption of Table 2: Do you mean "correlation coefficient", usually denoted by "R" or "coefficient of determination", is usually denoted by R2 ?

Coefficient of determination is correct; we changed the expression in the revised version.

p. 4761, Fig. 6: The bars in light grey are hardly readable. The authors should mind the basic considerations of colour contrasts.

Revised version is larger and has better contrast

Referee #2

{\it Title: One would more accurately say this paper has evaluated four instruments using three spectroscopic techniques rather than the techniques themselves.

We agree that the paper compares actually four instruments using three different techniques. In order to avoid a complicated and long title, we suggest following title: Evaluation of new laser spectrometer techniques for in-situ carbon monoxide measurements.

A sentence or two giving the analytical basis of each instrument would be valuable in later discussions. For example: 'The VURF is based on the fluorescence of CO at 150 nm (Gerbig et al, 1999). The discharge of a CO resonance lamp, excited by a RF discharge, is filtered and directed by two CAF2 lenses into the fluorescence chamber with a PMT.

We included a short description of the measurement principle for all for instruments in section 2.

P.5, line 6: Do the authors mean the raw signal was converted to mole fraction using a subjective sensitivity? Please clarify this.

The mole fraction was constant during this experiment. The raw data was converted to the same average mole fraction for all four analysers. In principle, this corresponds to a calibration which took place over the duration of the experiment.

P.5, Figure 2: Both QCL instruments show small enhancements at 2300 and at 0400 hr. Any ideas why these occurred?

Most likely small changes in the laboratory temperature or it could also be coincidence.

P.6, line 7, Figure 4: Are this daily averages?

The working standard was measured every 1-3 days for 15 minutes. The last 10 one minute values were averaged (blue open circles) and the error bars show the corresponding standard deviation expanded by a coverage factor k=2.
A sentence was added: 'The working standard was measured over a period of 15 minutes, and the last ten 1-min data were used for the calculation of the average and the standard deviation.' And in the figure caption: Each measurement (open blue circles) denotes to the average of ten 1-min values.'

P.6, line 15, Figure 3: It would be helpful if the plots showed the point chosen as the minimum Allan variation.

Red vertical lines were added to the plots indicating the optimal averaging time for each instrument.

P.6, lines 16-18: Is the degradation of the optics a process occurring in hours, days or years?

It depends on e.g. pollution and the quality of the operating gases; if polluted air is measured, the degradation is faster. Usually, cleaning of the optics is needed once per year if the analyser is continuously running. Added a sentence: 'Usually, cleaning of the optics is required once per year under normal operating conditions.'

Section 3.1.2: Temperature dependence - The authors should add a sentence explaining why these techniques are sensitive to temperature.

A sentence was added: 'Temperature changes potentially influence the measurements e.g. through line broadening, changes in laser intensities and detector sensitivities.'

P.6, line 31: 'for' should be 'four'

Done.

Section 3.1.3: Linearity: The authors should note that the linearity of a system is defined in part by the uncertainties of the measurement. Therefore a system can only be considered linear within its measurement error.

Yes, this is correct. We clarified '…were linear within the uncertainties of the experimental set-up'.

Figure 6 caption: 'not' should be 'note'.

Done.

P.8, line 15: Note the Nafion drier is an internal part of the instrument.

Changed to ‘The sample air of the VURF instrument is dried with a built-in Nafion® drier…’

P.9, Figure 8: The grey does not reproduce well. Choose another color.

Done.

P.9, line 17, Figure 9: The relationship between CO ratio and water vapor looks constant with any amount water vapor. The curve is somewhat deceiving. The water vapor effect needs additional comment.

It is correct that the ratios look constant for individual levels of CO; however, there is a clear difference between the lowest CO mole fraction and the other results. Since the instrument has already a correction implemented to account for the water vapour effect, one can see that this correction is not good enough for this particular instrument. Added a sentence: ‘However, the tested analyser showed a clear relationship between water vapour and CO mole fraction; low CO mole fractions tend to be overestimated in the presence of water.’

P.9, lines 15-20: Were experiments conducted where an external dryer was used? How do the water vapor corrections compare to water vapor free air samples?

First question: These experiments can only be done with humid air (or changing humidity). It is technically not possible to use an external drier. Second question: This information is available from the figures. If CO(reported)/CO(dry) is =1, then the correction is perfect.

P.10, lines 1-3: I am not sure what the authors mean here. Could it be re-written more clearly?

We also looked at the difference between CO(reported) and CO(dry) instead of the ratio; this information is actually not important and is not needed in the revised manuscript. We deleted these lines in the revised version.

P. 10, line 16: Was the CDRS instrument with the improved water vapor correction tested similarly to the first instrument for precision, linearity, etc? Performance of the improved model may be more useful to the reader.

Some tests were also made with the improved instrument. The results were similar except for the water vapour dependency (shown), and the instrument precision was also about a factor 2 better (not shown). However, we prefer to show the data of the original instrument, since all experiments were made simultaneously under comparable conditions. Furthermore, other analysers exist also in improved versions (e.g. enhanced performance version of the LGR instrument) which were not tested for the current study.

P.10, line 21: ‘implicates’ should be ‘indicates’

Done.

Section 3.1.5: Instrument summary. I suggest the authors remove all discussion of the FTIR instrument since its evaluation was not presented in sufficient detail.

We believe that the references to the FTIR technique are important, since it is the only new technique which has not been considered for this study. In order to give the reader an idea of the performance of this technique, we also feel that the data in table 2 should remain in our manuscript. If more detail on the instrument performance is required, the reader should refer to the cited FTIR papers.

P.11, line 12: This sentence should be written as ‘The instruments described above measured ambient air . . .’

Done.

P.11, lines 18-19, Figure 13. The top panel might show only the ICOS measurement as the bottom panel illustrates the differences among instruments much better than the overlay.

We decided to keep the original figure; the colors and the legend in the top panel are the same as in the lower panel.

P.11, lines 31-32: Instead of 'post calibrated' I suggest the text read '. . .were calibrated after the comparison based on. . .'

Done.

P.12, lines 21-24: Instead of '. . .raw data was in a first step..' this could be '. . .raw data was first. . .'. The following sentence '. . .in a second step adjusted to the working standards..' is unclear. This should be rewritten for better clarity?

Done.

P.12, line 29. 'Precision/repeatability' should be defined when first used in section 3.

We added a reference to the WMO/GAW Glossary of QA/QC-Related Terminology at the end of section 3.

P.13, lines 1-7: It should be noted that instrument drift can be accounted for with proper calibration.

Yes. We added '…and more frequent calibrations intervals would be required to account for this.'

P.13, lines: 27-29: This is not clear. Do the authors mean - 'This clearly indicates that the frequency and temporal averaging of the instrument signal cannot be neglected in locations of high short term CO change.'

Yes, this sounds better. We changed the sentence accordingly.

Table 1: I suggest using the terms 'stronger and weaker' rather than 'good or poor'. Remove the FTIR results.

We changed to weaker/stronger as suggested. FTIR results: see above.