

Referee #1

We would like to thank Referee #1 for the valuable comments and her/his time to review the manuscript. The questions/comments/concerns of Referee #1 (bold text) are addressed below. New text in the manuscript is given in italics.

General comment made by Referee #1:

The introduction section describes the background of the field and the aim of the paper shortly but still adequately. Results are dealt with high expertise and there is no doubt that authors have a deep understanding of the field. Since some graphical results are complicated, slightly longer explanations would help the reader, without a need to find a clarification from references. Tables and graphs itself are clear. Authors are referred to the key articles of this topic. The English of this paper is mainly sufficient. Altogether, my opinion is that the scientific quality of this manuscript is high, and it should be published with revisions.

Thank you for this comment. We agree that especially the bias / slope plots might be difficult to understand without further explanation. We refer to Zellweger et al. (2016) for a detailed description of the methodology, including an illustrative plot showing different bias / slope combinations. We believe that duplication of this explanation is not needed here. However, we added a sentence to clarify on page 6 / line 27: "*Perfect agreement would result in bias / slope pairs of (0 nmol/mol / 1)*".

Specific comments made by Referee #1:

-page 4, section 2.1: there is probably a SOP for the TS measurements at the stations? Like, how the TSs are prepared for the measurement? Is there used a similar sequence for TSs as the station uses for its own cylinder measurements or instructed by WCC? Is the protocol dependent on the instrumentation? Who is measuring the TSs in practice? I am not asking to answer all of these, but just shortly clarifying how the TS measurements done; with a certain fixed protocol or not; since this may affect to the results as well. Stations personnel may have bad habits, but they may also know by experience if their instrument needs special way to handle it.

Yes, there is a SOP for TS measurements at stations, which is available from the WCC-Empa webpage (<https://www.empa.ch/documents/56101/250799/2.pdf/f5a8c0a5-884f-4e0c-b836-96d92dbd260c>). However, there is no fixed protocol, since this depends on both the instrumentation and the capabilities of the stations. Usually, station operators are analysing the TS as unknown samples on their system, with at least three repetitions for each TS. It is clear that this approach has limitations, and additional parallel measurements with a fully independent analytical system are desirable and are implemented whenever possible during station audits. Brewer et al. (2019) recently showed that results of the audit approach by WCC-Empa are comparable to the WMO round robin experiments (NOAA, 2018). This confirms that the protocol followed by WCC-Empa is a valid approach to assess instrument performance. The limitations of the TS method are listed in section 2.2. We added the following sentence in section 2.1:

Usually, multiple analysis of a set of three or more TS is made and averaged for the final assignment of the TS value by the audited laboratory.

-page 5, lines 11-12: two questions concerning water vapour correction of WCC. 1) Since droplet method is not giving appropriate results, but in page 12 you are listing two - most probably superior - methods to determine water vapor correction function, why WCC is not using or even testing those

methods? 2) If the correction of the effect of water vapour to CO measurements is difficult, why WCC is not drying the sample?

1) It is true that the droplet method has limitations regarding the determination of the water vapour interference for CO. The methods described later in the paper were not yet published / available when the experiments done for this paper were made. Potentially, they would give better results, which would allow for a better compensation of the bias. However, a large part of the uncertainty is due to relatively large instrumental noise of the Near-IR-Cavity Ring Down Spectroscopy (NIR-CRDS) technique, which makes the determination of the appropriate correction function challenging. Furthermore, unpredictable short-term changes as shown in section 3.2 / Figure 10 further complicate appropriate correction.

2) As a consequence of the results, we are now drying the air, which is also one of the recommendations of this paper (see conclusions).

-page 8, lines 27-31/Figure 5: concept "relevant amount fraction" for N₂O remains a bit unclear. So, is the "relevant amount fraction range" a central 10 nmol/mol range representative for the unpolluted troposphere for the year of the audit and the "relevant amount fraction" a single representative value ±0.1 nmol/mol? Please try to state this more explicitly in the text.

Yes, this is correct. The amount fraction range is given in Table 2 for all audits. We agree that the description of the concept regarding the representative amount fraction was not clear in the manuscript. The relevant amount fraction is a single value and depends on the year of the audit. It was calculated using the value of the unpolluted air in 2016 and an annual growth rate of 0.8 nmol/mol. This value corresponds to the centre of the amount fraction range given in Table 2.

We made the following changes in the manuscript to better explain the concept:

Page 8, line 7: added "range".

Page 8, line 31: we added "*The relevant amount fraction corresponds to the value at the centre of the relevant range for the corresponding year.*"

-pages 8-9, section 3.1.2: it is clear how and why the compatibility goal for N₂O is determined. However, you clearly show in this manuscript that it will not happen in near future that atmospheric N₂O measurements will reach the goal. So, should the goal be revisited? I would like to see a statement in conclusions section regarding the necessity of these goals no-one is reaching.

We added the following sentences (from the report of the 19th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2017)) after the definition of the network compatibility goals in the introduction (page 2, line 20ff): "*These goals represent the maximum bias that can generally be tolerated in measurements of well-mixed background air used in global models to infer regional fluxes. Some network compatibility goals may not be currently achievable within current measurement and/or scale transfer uncertainties. However, they are targeted for applications which require the smallest possible bias among different datasets or data providers, such as for the detection of small trends and gradients (WMO, 2018).*"

We believe that it is beyond the scope of this manuscript to make recommendations regarding the compatibility goals, since this is done in the larger scientific community during the GGMT meetings.

-page 11, lines 1-2/Figure 10: if I understood correctly, the droplet tests are always made with the same CO amount fraction (300 nmol/mol), but is the water vapour response dependent on amount

fraction of CO, i.e. have you run the test with other CO amount fractions? It is a bit hard to see from the graph 10; there is no chronological trend, but the water vapour response is varying randomly?

We would like to thank the reviewer for this comment, since it made clear that our description of the experiment was not detailed enough. The water vapour experiments were made at different amount fractions ranging from 57 to 741 nmol/mol CO, but the limits shown in Figure 10 refer to an amount fraction of 300 nmol/mol. We found no dependency of the water vapour correction function on the amount fraction for both instruments, which is shown in Figure R1 where the ratios of CO(humid, corrected) / CO(dry) at a water level of 3% are plotted against the CO amount fractions used for the corresponding determination of the correction function. The limits for the WMO/GAW compatibility goal (green) and extended (yellow) compatibility goals are also shown.

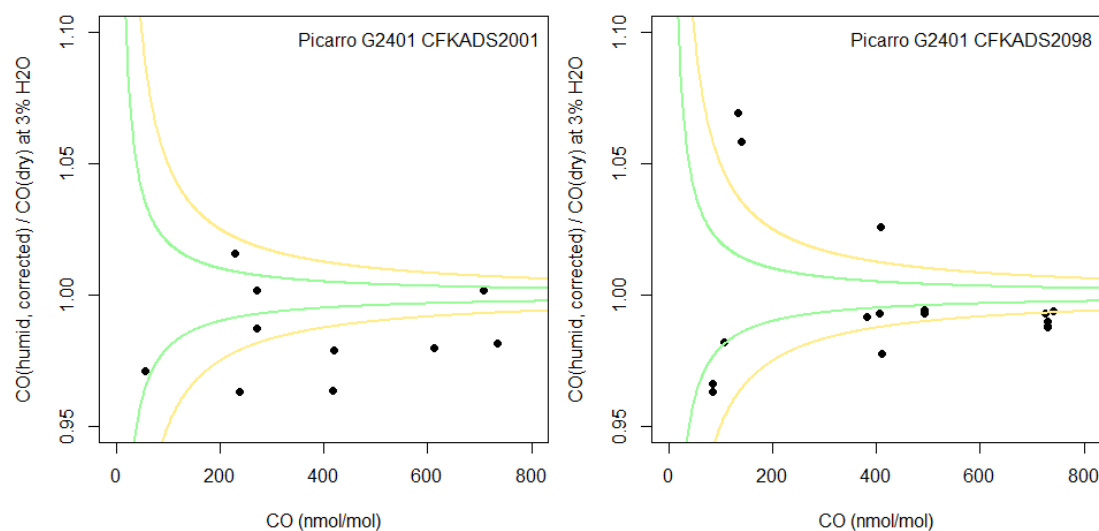


Figure R1. Black dots show the ratios of CO(humid, corrected) / CO(dry) amount fractions at a water level of 3% vs. the CO amount fraction of the CO standards used during the corresponding experiments of two different Picarro G2401 NIR-CRDS analysers. The coloured lines show the limits for the WMO/GAW compatibility goal (green) and extended (yellow) compatibility goal.

We also were not able to see a clear temporal trend in the change of the correction function, which is illustrated in Figure R2 below where the ratios of CO(humid, corrected) / CO(dry) at a water level of 3% are plotted against the time in weeks since the first determination of the correction function. The limits for the WMO/GAW compatibility goal (green) and extended (yellow) compatibility goals are also shown in Figure R2 as horizontal lines for an amount fraction of 300 nmol/mol CO. It looks like the Picarro G2401 CFKADS2098 gets less stable after approximately 150 weeks. However, both instruments were not able to fulfil the WMO/GAW network compatibility right from the beginning at a water level of 3%. This clearly supports the conclusion of our work that the internal water vapour compensation is inappropriate, and drying of the air sample is recommended.

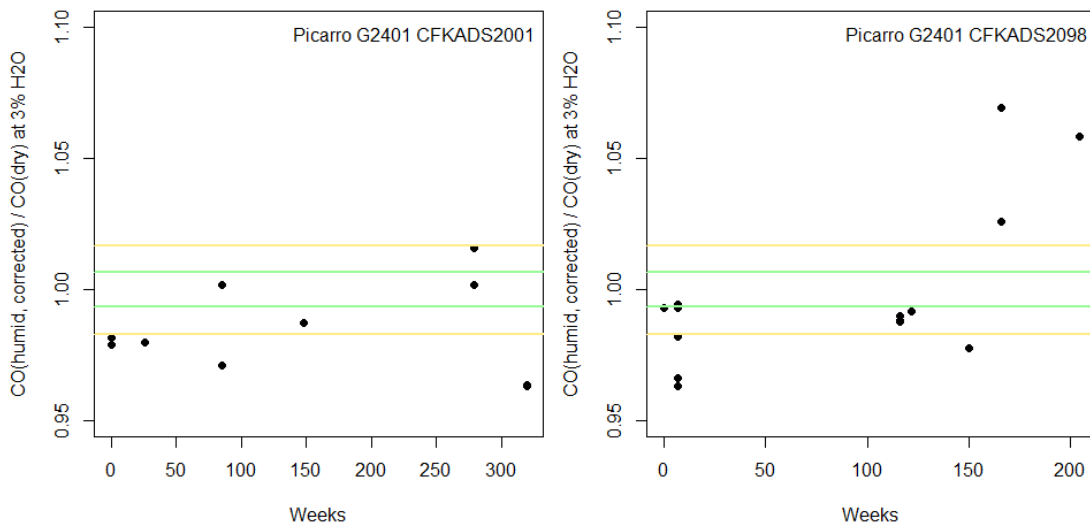


Figure R2. Black dots show the ratios of $\text{CO}(\text{humid, corrected}) / \text{CO}(\text{dry})$ amount fractions at a water level of 3% vs. time in weeks since the first determination of the water vapour correction function of the two different Picarro G2401 NIR-CRDS analysers. The coloured lines show the limits for the WMO/GAW compatibility goal (green) and extended (yellow) compatibility goal at an amount fraction of 300 nmol/mol.

To clarify the fact that different CO standards were used, we added a sentence in section 2.2: "The CO amount fraction of the standards used for the determination of the water vapour interference ranged from 57 to 741 nmol/mol. No dependency of the water vapour interference on the CO amount fraction was observed."

-page 11, lines 4-9: you state that the clearly decreased bias between TI and PUY instrument was thanks to the drying of sample air for TI. However, you also changed the inlet line and never tested the other line with dryer. Please, mention this uncertainty more clearly in the text.

Thank you for this comment. It is correct that changing the inlet line causes additional uncertainty. However, there is evidence that the improvement was due to the drying, since parallel measurements were also made for CH_4 and CO_2 over the same inlet lines, and no change in the bias was observed for these two compounds with and without drying. In addition, we have seen similar results during parallel measurements at other stations of the GAW network. For example, a recent parallel measurement was made at the GAW station Bukit Koto Tabang. In this case, the WCC-Empa TI was equipped with a Nafion dryer for the entire comparison period, but the station analyser was measuring humid air for the first week. The bias was significantly larger during the period when the station analyser was measuring humid air, and decreased after the installation of a Nafion dryer in the same inlet line, as shown in Figure R3.

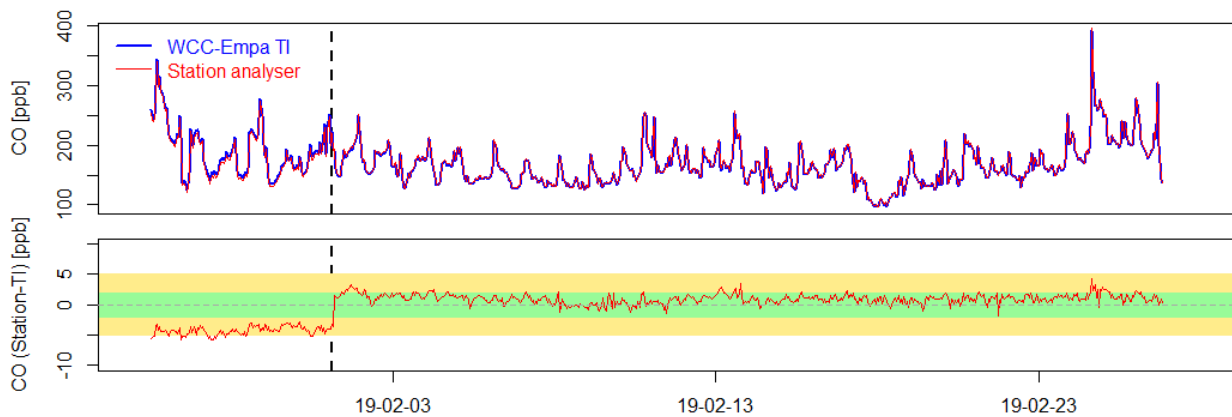


Figure R3. Comparison of the Bukit Koto Tabang Picarro G2401 analyser with the WCC-Empa travelling instrument for CO. Time series based on hourly data as well as the difference between the station instrument and the TI is shown. The coloured horizontal areas correspond to the WMO/GAW compatibility (green) and extended compatibility (yellow) goals. The dashed vertical line indicates the time of the installation of the Nafion dryer.

To clarify that the observed improvement is most likely due to the drying system but also to highlight the potential influence of the inlet system, we added the following sentences on page 11, line 8: "Potentially, the change of the inlet system could also have been the reason for the reduction in the bias. However, this is unlikely because no change in the bias of CH₄ and CO₂ amount fraction, which were both measured simultaneously together with CO over the same inlet line, was observed."

-page 11, line 13: did you use the same individual instrument (TI) in PYU and in AMY? Differences between Picarro G2401 individuals, especially regarding CO performance, may be high.

Yes, the same instrument was used at PUY and AMY (Picarro G2401 CFKADS2098) (see Table 3, instrument serial number is given there).

-page 11, lines 16-19: since there was three clearly different periods; due the calibrations of AMY instrument; during the comparison measurements at AMY, please provide the biases for each period separately.

We added the bias for the individual periods. We also realised that the bias of 0.23 ± 8.81 nmol/mol for the entire period given in the discussion paper was wrong, since it referred to a data set that included invalid data. We corrected to 0.10 ± 3.20 nmol/mol.

-page 12, line 18: is the drying of the sample air the only option, as you state, or is it possible to reach the goal with well and frequently made water vapour tests as well?

Because sudden changes of the water vapour interference were observed, a correction based on frequent determination of the correction function might still be insufficient. Furthermore, determination of individual correction functions are associated with large uncertainties and drying therefore is recommended.

You also leave open the questions how to dry and how low water levels are needed to reach. By looking the Figure 10, it looks that water levels below 0.3 %, roughly, would give CO values stable enough when Picarro G2401 instrument is used. That kind of a water levels are possible to reach by using Nafion dryer, for example. The drying issue is an endless story, but open this theme with a few sentences.

Yes, this is correct. It only is important that the water vapour level remains as constant as possible. Calibration gas and working standards also need to pass through the Nafion dryer to compensate for potential loss in the system.

We added the following sentences at the end of section 3.2:

"However, this will most likely be insufficient to detect the sudden changes in the correction function that were observed in our experiments. Consequently, drying of the sample air should be considered when measuring CO with a Picarro G2401 instrument. Both cryogenic traps and Nafion dryers can be used. WCC-Empa now uses Nafion dryers for the parallel measurements during station audits. Both single tube (MD-070-48S-4) and multi tube (PD-50T-12MPS) Nafion dryers in reflux mode using the Picarro pump for the vacuum in the purge air were successfully used. This reduced the amount of water to approximately 0.06 – 0.22% (single tube) and 0.01-0.03% (multi tube), depending on ambient air humidity. In case of using Nafion dryers, the standard gases must also pass through the dryer to compensate for a potential loss over the dryer."

And in the conclusions:

"This can be implemented through drying of the sample air with cryogenic traps or Nafion dryers."

-figure 7: since the water vapour was probably the main reason for bad agreement between TI and PUY instrument, please add a panel showing the H₂O % reading of TI.

The reason for the bias is the inappropriate compensation of the water vapour interference made by the Picarro instrument. There is a dependency of the bias on the water vapour level, which we showed in Figure 8. Figure R4 shows the original Figure 7 of the paper with an additional panel showing the H₂O reading of the instrument, as requested by the referee. However, we feel that this provides only minor additional information compared to the dependency on the water level shown in Figure 8. We therefore prefer to keep the figure as it is.

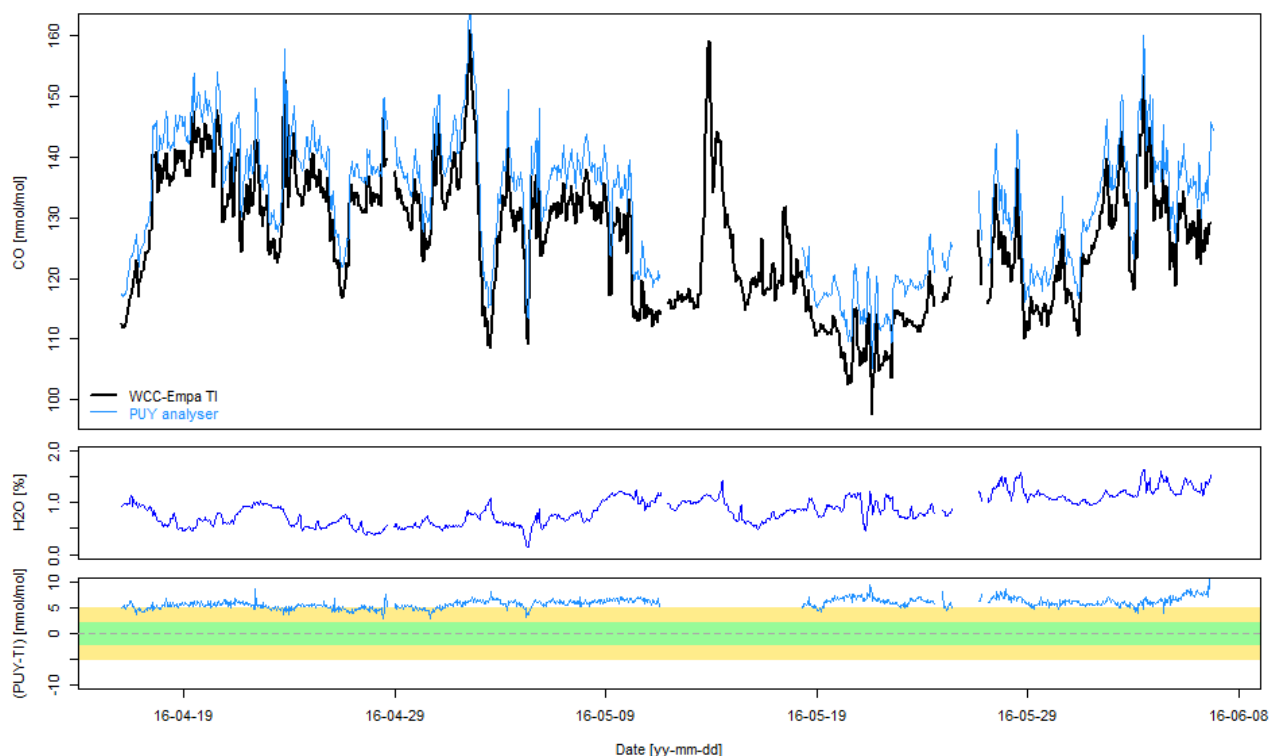


Figure R4. CO comparison at PUY between the WCC-Empa travelling instrument and the PUY Picarro G2401 for the period when the TI sampled humid air. Upper panel: CO time series (1 h data). Middle panel: H₂O measurements of the TI. Lower panel: CO bias of the station analyser vs time. The green and yellow areas correspond to the WMO compatibility and extended compatibility goals.

-figure 13: please, mark the calibrations of AMY instrument in lower panel with vertical lines, for example.

We added vertical lines in the revised version of the paper.

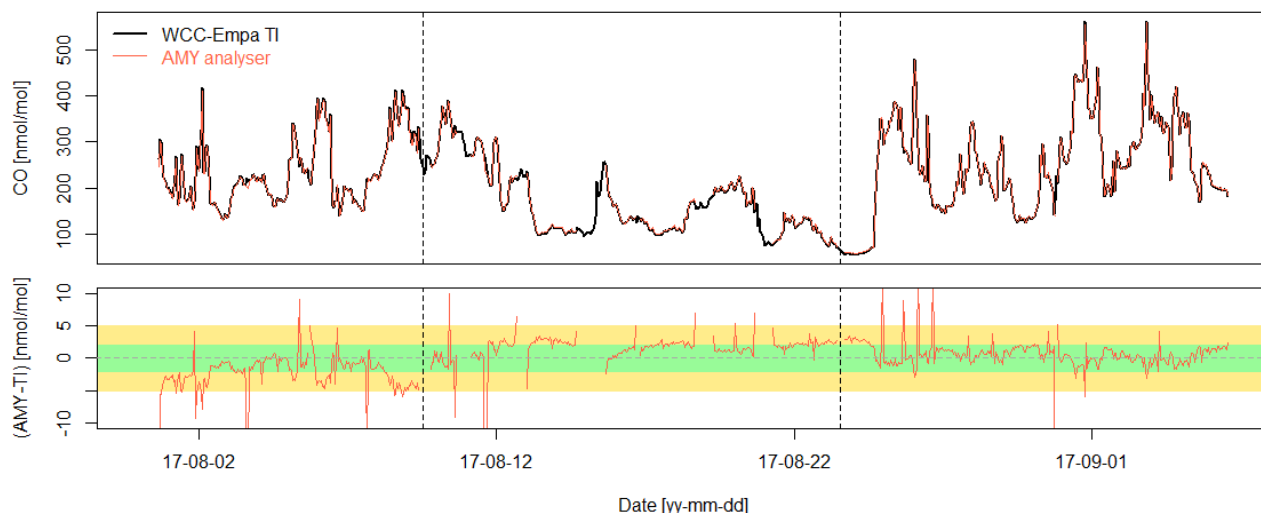


Figure 1. CO comparison at AMY between the WCC-Empa travelling instrument and the AMY Los Gatos 30-EP QCL analyser. Both instruments sampled dry ambient air. Upper panel: CO time series (1 h data). Lower panel: CO bias of the station analyser vs time. The green and yellow areas correspond to the WMO compatibility and extended compatibility goals. The dashed vertical lines indicate the time of the calibration of the AMY instrument.

Technical corrections suggested by Referee #1:

-page 1, line 20: "analyse" > "analysed".

Accepted and changed.

-page 2, line 1: "GAW programme" is written with and without a capital "p" in this manuscript. Please uniform using lower case "p".

Accepted and changed throughout the paper.

-page 2, line 11: What is "Empa"? Please add institute full name and location.

Done.

-page 2, line 11: Use "WCC" instead of "World Calibration Centre". There is often used the full name instead of abbreviation, even though the abbreviation is introduced when first time mentioned. Like three lines below "World Calibration Centre" in written again, and with gas components and instrument techniques is the same issue. Uniform all these issues in the whole manuscript.

Done.

-page 2, line 14: WCC-N₂O, who operates? It is written on page 3, so move it to here.

Done.

-page 2, lines 28-30: introduction of the GC detectors for CO; now it is not fully clear that CG/FID with methaniser and CG/HgO are two separate instruments to measure CO. Please reword.

We reworded the sentence to "..., whereas flame ionization detection (GC/FID) in combination with a methaniser and GC with mercuric oxide reduction detector (GC/HgO) *were the two most commonly used techniques for CO measurements* (Zellweger et al., 2009)."

-page 4, lines 9-10: move the operator of CCL to where it is first time mentioned.

Done.

-page 5, line 6: add comma after "measurements".

Done.

-page 6, line 15: maybe "trend in the atmospheric concentrations" over "in the atmosphere".

Changed to "... due to the significant upward trend of *the N₂O mixing ratio* in the atmosphere..."

-page 7, lines 13-14: what is the meaning of "(see also Figure, right)"?

That was a mistake and we deleted it.

References

Brewer, P. J., Kim, J. S., Lee, S., Tarasova, O. A., Viallon, J., Flores, E., Wielgosz, R. I., Shimosaka, T., Assonov, S., Allison, C. E., van der Veen, A. M. H., Hall, B., Crotwell, A. M., Rhoderick, G. C., Hodges, J. T., Mohn, J., Zellweger, C., Moossen, H., Ebert, V., and Griffith, D. W. T.: Advances in reference materials and measurement techniques for greenhouse gas atmospheric observations, *Metrologia*, 56, 034006, 2019.

NOAA: WMO/IAEA Round Robin Comparison Experiment, <https://www.esrl.noaa.gov/gmd/ccgg/wmorr/index.html>, last access: 23 July, 2018.

WMO: 19th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2017), Dübendorf, Switzerland, 27-31 August 2017, GAW Report No. 242, World Meteorological Organization, Geneva, Switzerland, 2018.

Zellweger, C., Emmenegger, L., Firdaus, M., Hatakka, J., Heimann, M., Kozlova, E., Spain, T. G., Steinbacher, M., van der Schoot, M. V., and Buchmann, B.: Assessment of recent advances in measurement techniques for atmospheric carbon dioxide and methane observations, *Atmos. Meas. Tech.*, 9, 4737-4757, 2016.

Zellweger, C., Hüglin, C., Klausen, J., Steinbacher, M., Vollmer, M., and Buchmann, B.: Inter-comparison of four different carbon monoxide measurement techniques and evaluation of the long-term carbon monoxide time series of Jungfraujoch, *Atmos. Chem. Phys.*, 9, 3491-3503, 2009.