We thank the three anonymous referees for their comments on our paper titled "Evaluating the performance of commonly used gas analysers for methane eddy covariance flux measurements: the InGOS inter-comparison field experiment". The comments are addressed individually below. The referee comments are shown with bold font and our responses are shown as plain text.

Anonymous Referee #1

<General comments>

In this manuscript, the authors present an inter-comparison of eight fast and precise CH_4 gas analysers suitable for eddy-covariance flux measurements using a dedicated and well-designed 3-weeks experiment. This methodological and technical topic is useful for the scientific community since the number of ecosystem-scale measurement sites of CH_4 fluxes is growing fast with the aim to better constrain the CH_4 global cycle. Reliable flux measurements are mandatory in this context and gas analysers, which are rapidly improving, are the cornerstone of these systems.

The paper shows some similarities with a previous and recent publication by the same first author (Peltola, O. et al., 2013, "Field inter-comparison of four methane gas analyzers suitable for eddy covariance flux measurements." Biogeosciences 10(6): 3749- 3765), already comparing four of these height analysers but it now includes latest models and an original part of the present paper is the thorough discussion about the potential artifacts that can occur when correcting CH_4 fluxes from H_2O interferences. The magnitude of these effects is site and set-up specific and cannot therefore be extrapolated from this paper to all situations but the proposed methodology to handle the problem will be useful for the eddy-covariance CH_4 community, especially for non-experts dealing with slightly older analyser models.

The experimental set-up was not always optimal for testing the bulk performances of the analysers (separate intake tubes with different flowrates, logging problems with the LI-7000 and the G2311-f) but these limitations are properly dealt with in the discussion. Different plumbing schemes have also the advantage to allow testing robustness of the correction procedures. One missing point is that the authors should recognize explicitly that this short campaign was not relevant for testing the long-term behavior of these analysers (long-term stability, clogging of the cell, easiness of maintenance,…) which is an important aspect when choosing an apparatus for long-term campaigns.

State-of-the-art measuring techniques were used. The writing style is excellent, almost no typos, figures are clear and well introduced and reference to literature is appropriate and complete. I therefore recommend this paper for publication provided the minor (mainly technical) comments below are considered.

We thank the referee for acknowledging our work. All the referees mention that this campaign was not long enough for assessing long term behaviour of the tested analysers. We agree with this notion and will bring this forward more clearly in the revised manuscript.

<Specific comments>

P800L4: Precise what do you mean by “important”. Since H2O is mentioned, I guess you mean here through its role in the radiative balance of the atmosphere (and not through its forcing effect which is evoked in the next sentence).

The first sentence of the Introduction section will be replaced with “Methane (CH4) is the third most important greenhouse gas for the radiative balance of the atmosphere, after water (H2O) and carbon dioxide (CO2).”

P803L25: You should use micromol mol-1 instead of ppm

Will be replaced.

P807L22 and P808L1-4: Which kind of model was used for fitting the ensemble averaged temperature cospectra? Why didn’t you use the traditional Kaimal’s parameterization, which should be well verified on such an ideal site?

The referee is right; usually at these kind of flat and level surroundings the Kaimal’s parameterisation should work well. However, when plotting the data we noticed that there was a small, but clearly detectable difference between Kaimal’s parameterisation and our measured temperature cospectra. Especially the location of cospectral maxima did not follow Kaimal parameterisation. Thus we decided to fit our own cospectral model.

In unstable cases we used

$$\begin{cases} \dfrac{fC_{w\theta}}{\overline{w'\theta'}} = \dfrac{24.9n}{(1+48.5n)^{1.4}}, n<1 \\ \dfrac{fC_{w\theta}}{\overline{w'\theta'}} = \dfrac{3.5n}{(1+3.8n)^{7/3}}, n\geq 1 \end{cases}$$

where f is natural frequency in Hz, $C_{w\theta}$ is cospectral density of the vertical wind component (w) and temperature (θ) and n is normalised frequency ($n = f(z-d)/U$, where z is measurement height, d is displacement height and U is wind speed). In stable cases we used

$$\begin{cases} \dfrac{fC_{w\theta}}{\overline{w'\theta'}} = \dfrac{6.2n}{(1+7.2n)^{1.8}}, n<1 \\ \dfrac{fC_{w\theta}}{\overline{w'\theta'}} = \dfrac{8.2n}{(1+3.8n)^{2.6}}, n\geq 1 \end{cases}$$

Furthermore, the location of cospectral maxima (n_m) was parameterised based on the obtained data. In unstable situations it was set to constant with a value of 0.07, whereas in stable situations it followed

$$n_m = 0.07\left(1+5.7\left(\frac{z-d}{L}\right)^{0.5}\right)$$

where L is the Obukhov-length.

P827L1: Please consider adding a figure to illustrate your observed bias in the CH4 flux measured daily cycle in case of faulty flux computation. The question of the robustness of measured daily cycles is an important one in the community and I think it's worth a figure.

We will add a figure illustrating the bias in the daily cycle of CH4 flux due to incorrect flux calculation. The following sentences discussing the new figure will be added to page 822 line 18: "Systematic bias in the daily cycle of FGGA CH4 flux in these four cases is shown in Fig. 12. The daytime CH4 fluxes are overestimated in the "no revision" case (Fig. 12a) and in the "attenuation revised" case (Fig. 12b). This is in line with Fig. 11a and 11b, since at daytime the latent heat flux is high and according to Fig. 11a and 11b this increases also the bias in the correction. In the "lag time revised" case no clear bias in the daily cycle can be seen, whereas in the "both revised" case daytime CH4 fluxes are underestimated." The figure will be referred to also on page 827.

<Technical comments>

P801L6: Typo: "main" instead of "mains".

"Mains power" is correct English in this sentence.

P804L22: Typo: a verb is missing.

The sentence will be reformulated and replaced with "Air for five closed-path gas analysers was sampled from near METEK2."

P806L22: Add the definition of molar mixing ratio here: "of gas c (ratio of gas c mole number to those of dry air)" and use latter on only "mixing ratio" and not "dry mixing ratio" like on L24.

The nomenclature will be revised everywhere in the text as follows: dry mole fraction will be used for r_c (previously it was referred to as molar mixing ratio) and wet mole fraction will be used for χ_c (previously it was referred to as mole fraction). This is closer to the nomenclature used in ICOS community and the one recommended by IUPAC (International Union of Pure and Applied Chemistry). Definition of dry mole fraction will be added to page 805, line 22: "($r_c = \frac{N_c}{N_{\text{dry air}}}$, where N is number of moles)."

P809L19: Insert this definition of "dilution" after the first occurrence of this term on L16.

The definition of dilution will be moved.

P814L4: Typo: remove the first "values"

Will be removed.

P818L10-18: Consider moving this paragraph to P810L20, in the MM section, and shorten it. This is just a confirmation of literature.

We disagree that this is just confirmation of literature since Eq. (5) has not been presented in the literature before our study. Thus it is good to show that the two correction methods (Eq. (4) and (5)) give the same end result, just as they should based on theory. A similar comparison has been done for the dilution correction in Ibrom et al. (2007) but they did not include the spectroscopic correction in their comparison as we do.

P822L1: "with the correction calculated with the internal H2O measurements and internal CH4 lag-time".

Will be added.

P824L21-P825L23: Move these two paragraphs to the introduction section.

We would like to keep these two paragraphs here, since here they first quickly introduce what kind of results has been presented in other similar studies before our results are discussed. This way they give a comparative background for the discussion.

P828L8: "of systematic bias".

Will be added.

P829L11: You can group the two identical terms in Equ. A1.

The two terms will be grouped into one.

Table1: Please unify the order of presentation of analysers in all your tables. You should group them following a systematic logic (group same manufacturer and/or techniques). Also in figures where it's relevant.

Gas analysers in tables and figures will be grouped based on manufacturer.

Figures: Consider removing all grids in your figures when they are not really useful.

Grids will be removed.

Figure3: Typo: "to medium" instead of "to mediocre"

Will be corrected.

Figure4: Second sentence hard to read; please split it in two parts.

Will be modified.

Figure10: The discussion linked to this figure is complicated (even if it's very precisely written and if I understood clearly your point). To lighten the figure, I would remove CH4 FGGA curve (its exact shape has no importance in the discussion) and add only a vertical line giving the lag time maximising its covariance. If you keep this curve, precise how it was normalize (by its maximum value). I would also avoid using the same colours in fig 11 since the figures are linked but the colours have different meanings in the two figures.

CH_4 FGGA curve will be replaced with a vertical dotted line and the colours in this figure will be changed.

Anonymous Referee #2

<General comments>

This paper presents the results from a two weeks field test of eight fast response methane analysers that were used in an eddy covariance (EC) setup in order to calculate the turbulent exchange of methane between a grassland site and the atmosphere. The data analysis covers the instruments' performance in terms of precision, necessary spectral and density corrections and consistency of calculated fluxes. The manuscript is very well written, the high quality data are thoroughly analysed and despite its complexity the paper is well structured and thus easy to read and to understand.

The available techniques to measure atmospheric methane concentrations at high frequencies – and the number of available sensors – have developed rapidly during the last years. Sharing the results from a systematic field comparison between the most promising of them is therefore highly relevant and useful for the scientific community. A particularly useful and novel aspect is the detailed analysis of the necessary density corrections due to simultaneous water vapour fluxes, which is not straightforward if the water concentrations are not measured in the same cell as the methane concentrations are. This is the case for some of the older instrument versions, and the corresponding uncertainties and biases have often been neglected in earlier EC studies.

Therefore I consider the manuscript as a suitable contribution to "Biogeosciences" and I expect that it has the potential to become a much cited reference paper. A problem I have with the paper is nevertheless the short duration of the field experiment that does not exactly allow wide-ranging conclusions about the sensors' performance in continuous EC studies. Therefore I recommend that the authors be encouraged to make a few clarifying changes to the text and to at least one table before a decision about publication in "Biogeosciences" can be made. The recommended minor revision should refer to the following specific comments and questions

Similarly to Referee #1, also Referee #2 pointed out that based on this study alone, no definite conclusions on sensors' long-term performance can be made. This will be brought forward more clearly in the revised manuscript.

<Specific comments>

Page 800/801: The second paragraph of the Introduction section raises the issue of long-term measurements at remote sites and the need to assess the suitability of gas analysers for such purposes. However, this question is not really answered in this paper because it contains only a very short data record (< 1 month) – in contrast to an earlier Peltola et al. (2013) paper in BG. The suitability of the analysers will actually depend on their stability, for example in terms of CRD time or corresponding indices, and the required maintenance work (filters, mirrors etc.). For instance, the necessity to clean mirrors every second day (section 4.1) would rule such an instrument out for certain setups. Please make absolutely clear what kind of conclusions can be expected from this paper and which questions it can NOT answer.

We will address this comment by adding the following sentence to page 801 line 21: "Due to its relatively short duration, our study cannot answer the question of how applicable the tested instruments are for long-term field usage, for that the reader is referred to Peltola et al. (2013) and Detto et al. (2011).". Furthermore, this will be brought forward also at page 825 line 23 with a sentence: "However, due to the fairly short length of our study, we cannot conclude which one of the tested analysers is the most suitable for long-term field application, for that the reader is referred to Peltola et al. (2013) and Detto et al. (2011)."

Page 804: Please explain whether the changes in the measurement frequency of the DLT-100 and the FMA2 were done accidentally or on purpose.

The measurement frequencies of these two instruments were accidentally too low during some periods of the campaign. The fact that it was an accident will be mentioned in the revised manuscript.

Page 814 bottom: How was the unusual threshold of u_star > 0.08 m/s identified? Please explain whether this was an arbitrary choice or a decision based on data analysis.

The threshold was based on data obtained during the campaign (see Fig. 1). This will be mentioned in the revised manuscript. Similar threshold value (0.07 m s^{-1}) was found by Kroon et al (2007) who presented CH_4 fluxes measured a few kilometres from our study site, in a similar landscape.

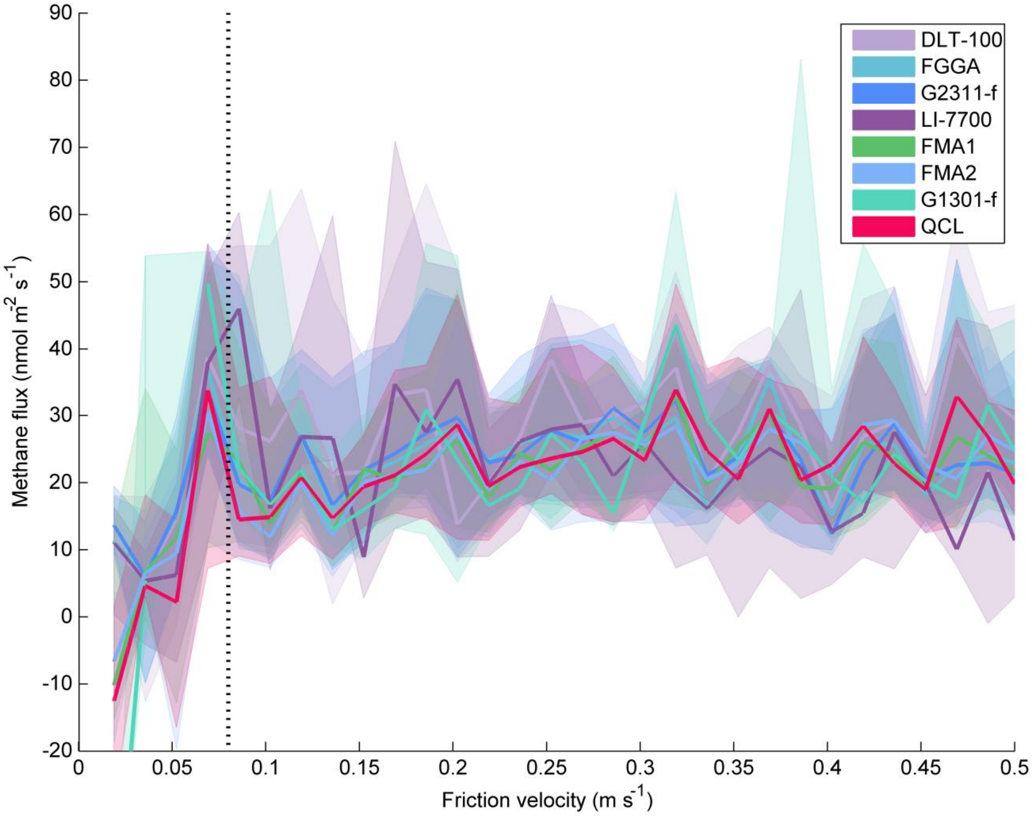

Figure 1: Methane flux from the tested instruments plotted against friction velocity. Data was grouped in friction velocity bins before plotting. The lines show medians and the areas show the interquartile range in each bin. The dotted line shows the friction velocity threshold used in discarding flux data.

Page 816 bottom: Why is the observation of a missing influence of CRD time on the noise "surprising"? I would leave this word out. As long as CRD time is above a certain threshold it will no longer affect the noise which would instead be dependent on other limiting factors such as the sensor's general resolution limit or pressure fluctuations in the cell.

Due to this comment, word "Surprisingly" at page 816, line 28 will be replaced with "However"

Page 817 bottom: Was the low measurement frequency, the resulting artificial (linearly interpolated) 10 Hz data and the corresponding slow response time of the G2311-f (Fig. 6) in fact the reason for the - apparently - very low noise in the "raw" methane concentrations (Fig. 5) of this sensor? This should be discussed on page 825, line 25.

This is a good point which was not considered before. A few sentences about this issue will be added on page 816, line 12 onwards: "The low noise estimate for G2311-f data may be partly caused by the instrument's low measurement frequency (2.3 Hz; see Sect. 4.3.1 for explanation), because noise decreases with increasing the sample size. Noise in 10 Hz G2311-f data can be roughly estimated by multiplying the noise estimate for 2.3 Hz data (0.4 ppb) with $\sqrt{10Hz/2.3Hz}$. This calculation yields a value of 0.8 ppb, which is closer to the value estimated for G1301-f, but still a lot smaller than what was approximated for other instruments." Also the sentence starting from line 24 at page 825 will be reformulated as: "It was shown in Section 4.2 that the Picarro instruments G2311-f and G1301-f are superior to the other instruments tested in terms of signal noise, although the low noise estimate for G2311-f data could be partially explained by the instrument's low measurement frequency during our experiment."

Page 825: The discussion on long-term applications of the sensors is not sufficient. It appears as if data gaps, for example caused by rain for open-path analysers, would be the only issue with respect to an analyser's suitability. Actually, this short-term study cannot judge about the analysers' long-term performance. Please add a sentence or two about what else we would need to know to make a statement about which analyser(s) to choose for long-term applications.

The following sentences will be added at page 825, line 23 onwards: "Applicability of an analyser for long-term studies is also hampered if constant maintenance is needed (filter replacement, liquid nitrogen is needed for cooling the laser and/or detector, laser tuning, cleaning of measurement cell mirrors etc.), analyser drifts significantly or if the analyser malfunctions frequently and needs to be taken away from the measurement site for repair. However, due to the fairly short length of our study, we cannot conclude which one of the tested analysers is the most suitable for long-term field application, for that the reader is referred to Peltola et al. (2013) and Detto et al. (2011)."

Table 1 (and corresponding text on page 804): Here I am missing some very important information about the setup: What was the cell pressure in the analysers? Were the vacuum pumps able to keep it constant or did it vary during the experiment (and if yes, how much)? Did the pumps (Edwards vs. Varian) in combination with the respective filter settings perform equally well? This is important because the noise and the magnitude of the measured methane concentrations will depend on how well the target pressure in the cells could be maintained.

Analyser cell pressure will be added to Table 1, in addition to estimates of how much they were varying during the campaign. The following sentences will be added to page 804, line 21: "The cell pressure of DLT-100 varied slightly during the campaign, whereas FGGA had a relatively constant cell pressure throughout the measurement period (Table 1). Picarro instruments control cavity pressure rigorously by opening/closing valves in front of or behind the cavity, which keeps the cavity pressure practically constant." And to page 805 line 13: "Due to precise pressure control, the G1301-f cell pressure remained constant throughout the campaign. Cell pressures of FMA1, FMA2 and LI-7000 varied slightly and QCL cell pressure was not recorded (Table 1). No significant difference in performance between the different pump types used was found."

Looking at Fig. 1 I wonder why there were two additional sonic anemometers mounted on the tower – did they have any connection to the experiment described here?

After this campaign a second campaign was held where the CH_4 flux variability in the surrounding agricultural landscape was studied. These two additional anemometers were used in that study and were deployed during this first campaign only for validation purposes.

Anonymous Referee #3

<General comments>

The manuscript of Peltola et al. evaluates the performance of eight types of fast methane analyzers when used for measurements of methane fluxes by the eddy covariance methodology. The comprehensive study is very relevant since the analysis of methane fluxes between ecosystems and the atmosphere is a strongly growing research field. This intensification of methane research is fostered by the fast development of several competing lines of fast analyzers. As these methane analyzers are comparatively fresh on the market and have quite different technical requirements than the previously more extensively studied carbon dioxide/water vapor analyzers, methodological research is essential for establishing a sound basis of methane flux research. The manuscript by Peltola et al. is particularly useful for non-specialist scientists by making aware of - and giving solutions for - a wide variety of practical problems that are typically encountered when working with eddy covariance methane flux measurements. The derivation and discussion of equation (5) for correcting density and spectroscopic effects on half-hourly averaged fluxes is very interesting and highly valuable especially for users of analyzers that do not measure water vapor in parallel to the methane concentrations. I would recommend this paper to all scientists starting with eddy covariance methane flux measurements.

Unfortunately, the comparison of the only open-path instrument (LI-7700) with the closed-path instruments is hampered by an inappropriate data-logging set-up. Due to the degradation of the measurement resolution to about 20 ppb due to the data-logging set-up, a reasonable assessment of data quality and coverage is not possible. This problem is addressed by the authors at several places in the text; however, I think that this problem should be high-lighted even more carefully, also in the captions of Table 2, Table 5, Figure 3 and Figure 6. These comparative tables and figures are the easiest way to find information about how the analyzers' performances compared to each other. To avoid misunderstandings, the data logging problem should be mentioned directly at all of these important comparative tables and figures. It should be also made clear that this problem is an external and not an instrument-specific problem. (E.g., page 844, caption of Fig. 5: I suggest writing something like: "The LI-7700 data are clearly affected by an external data logging problem which is not caused by the analyser."

I agree with Referee #2 that the authors should better discuss the limitations of their study regarding the evaluation of analyzer suitability for long-term measurements. For long-term measurements, long-term stability is an important analyzer property which could not be investigated in the two week campaign presented by this manuscript.

The manuscript is very well written. I have only a few suggestions for setting some more commas and hyphens, which I list in the technical comments.

I recommend the manuscript of Peltola et al. for publication in Biogeosciences after minor revisions.

We thank the referee for recognizing our work. As the referee points out, it was really unfortunate that the LI-7700 data was degraded by the data logging system. The referee suggests that it should be brought forward more in the manuscript and we will follow his/hers suggestions. We will mention the data logging problem at all captions that the referee suggests and make clear that the degradation was caused by an external problem, not by the instrument itself. As mentioned earlier, the fact that our study cannot evaluate analysers' long-term field performance will be brought forward in the Introduction (Sect. 1) and Discussion (Sect. 5) sections.

<Specific comments>

Page 802, line 14: What kind of vegetation? Grass?

The vegetation consisted mainly of grass, but occasional sedges were growing next to the drainage ditches. This will be mentioned in the revised manuscript.

Page 802, line 18: Was this wind direction sector filtered out? Was the disturbance of this tall building considered problematic when wind came from this direction? Or is it assumed that problematic effects would be identified by the quality control tests?

Data from this wind direction was not omitted as it was assumed that the quality control tests were able to detect and remove faulty data.

Page 803, line 3 and line 15: Why was the reason to use two anemometers instead of one? Distance between sample intakes and anemometer?

When the campaign was planned it was estimated that sampling with eight closed-path gas analysers from one location would cause too strong suction in that location and thus would affect the flow field around the anemometer. Therefore we decided to split the analysers between two identical anemometers, in order to minimize the effect of sampling on the turbulence around the wind measurements. Distances between the sample intakes and corresponding anemometers are given in Table 1 and the distance between the two anemometers is given in Fig. 1.

Page 811, line 7: I recommend to mention already here that it has to be assumed that the drier completely removes the water or at least the water fluctuations.

Sentence starting from page 811, line 5 will be modified into "CH_4 signals from the G2311-f and the QCL were free from H_2O interference since the G2311-f applied a similar correction internally during the measurements and the QCL was connected to a drier and it is assumed that the drier completely removes the effect of water on the sampled CH_4".

Page 811, line 15: Another recent publication on the issue is Runkle et al (2012) [Runkle B.R.K., Wille C., Gažovič M., Kutzbach L. (2012): Attenuation correction procedures for water vapour fluxes from closed-path eddy-covariance systems. Boundary-Layer Meteorology 142(3): 401-423.]

Citation to this publication will be added.

Page 817, lines 10-12: Is there an idea why the FMA2 should have this structured noise?

No clear reason for this was found, however it was noticed that the high detection limit values for FMA2 occurred only after the instrument was brought back to the field after the cleaning operation on 19th of June. Before the cleaning operation the detection limit values were similar to what for instance FMA1 had (approximately 2 nmol m^{-2} s^{-1}). The following sentence will be added to the end of section 4.2: "However, before the instrument was taken back to the laboratory for cleaning (19th of June), the instrumental noise values were similar to those reported by other instruments (approximately 2 nmol m^{-2} s^{-1}), which suggests that cleaning of the cavity was unsuccessful."

Page 817, lines 25-26: Was this linear interpolation internally done in the G2311-f instrument?

Yes. The instrument saved two sets of files: the so-called "sync files", which contain both the anemometer data and G2311-f data, and "not sync files", which contain only G2311-f data. While the "sync files" contain 10 Hz data, the "not sync files" contain only approximately 2.3 Hz data. By comparing these two sets of files it was noticed that the 10 Hz concentration data in "sync files" were a linear interpolation of lower frequency data saved in the "not sync files".

Page 818, line 15: I suggest adding "nearly" before "identical"

Will be added.

<Technical Comments>

Throughout the text: Please write consistently "CH4" or "methane"

In the revised manuscript "CH_4" will be used.

Page 801, line 2: I suggest hyphenating "time-consuming"

Will be corrected.

Page 803, line 19: I suggest "different" instead of "separate"

Will be corrected.

Page 803, line 25: Insert comma after "Unfortunately"

Will be added.

Page 803, line 26: Insert comma before "and".

Will be added.

Page 803, line 28: Place comma after "oversight"

Will be added.

Page 804, line 19: Remove "measured" or "reported"; place comma before "and" “measured” will be removed and comma will be added.

Page 808, line 6: I suggest hyphenating: “system-specific”

Will be modified.

Page 808, line 13: I suggest writing “accuracy” instead of “precision” here.

“precision” in this sentence will be replaced with “accuracy”.

Page 808, line 19: I suggest adding “mole fraction” after “CH4”

Will be added.

Page 811, line 20: I suggest hyphenating: “sampling line-specific”

Will be modified.

Page 811, line 22: place comma after “however”.

Will be added.

Page 813, line 8: Place comma before “and”

Will be added.

Page 814, line 14: Remove “values” before “cross-covariance values”

Will be removed.

Page 816, line 29: Place comma before “and”

Will be added.

Page 817, line 1: Place comma after “For example”

Will be added.

Page 823, line 8: Place comma after “instruments

Will be added.

Page 839, caption of Table 5: I suggest starting a new sentence after “analysis”. As it is now, it reads awkward.

A new sentence will be started after the word “analysis”.

References

Kroon, P. S., Schrier-Uijl, A. P., Hensen, A., Veenendaal, E. M. and Jonker H. J. J.: Annual balances of CH_4 and N_2O from a managed fen meadow using eddy covariance flux measurements, Eur. J. Soil Sci., 61, 773-784, 2010.

Ibrom, A., Dellwik, E., Larsen, S. E. and Pilegaard, K.: On the use of the Webb-Pearman-Leuning theory for closed-path eddy correlation measurements, Tellus Ser. B-Chem. Phys. Meteorol., 59, 937-946, 10.1111/j.1600-0889.2007.00311.x, 2007.