Response to Reviewer #2's Comments

We thank the reviewer for his/her constructive comments. We addressed all of them and modified our manuscript accordingly. Detail answers are given below.

Reviewer #2 comments:

The study by Nakayama et al. describes a MS-based measurement system that allows the quantification of mixing ratios of the (trace) gases CO_2*,* CH_4*,* N_2O *and* O_2*. This is a novel approach, and certainly qualifies for a high-ranking journal like AMT. The authors have conducted both laboratory incubations and field measurements. For both applications, data evaluation is limited to* CO_2 *and* N_2O *since instrument precision prevents flux calculation for* CH_4 *and* O_2*. This is a pity, especially since the authors are*

i) suggestive of preparing another publication that shows CH_4 *flux rates determined in the field*
ii) provide the perspective that another data evaluation algorithm could increase precision for O_2 *measurements so that flux rates could be calculated.*

For a proof-of-concept study, this is a surprising selection of datasets and from studying the manuscript, I could not find the reason why the suggested waveform averaging was not applied instead of ion counting for O_2*. If this is not possible with the given data I have missed the explanation. For a journal like AMT, the advance for – in this case – determination of soil-atmosphere (trace) gas exchange has to be shown. Even though the manuscript describes progress in making high resolution mass spectrometry field deployable, the benefit of this approach compared to other existing methods is not evident at the moment because*

There are many instruments available for determination of CO_2 *at low cost.* CH_4 *and* N_2O *exchange between soil and atmosphere has been determined using gas chromatography and, more recently, using spectroscopic methods. With regard to the greenhouse gases, the presented method doesn't seem to reach the precision of existing methods, but the draft stops short of an actual discussion of precision levels (which could for example start with an assessment what precision levels are required for the intended application) for different trace gases and a comparison with available commercial products. For example, determination of the sink strength of upland soils for methane is a challenge at the moment.*

- Our mass spectrometry (MS) based soil gas measurement technique is a novel and a new addition for *multiple soil gas field measurements* and thus submitted to AMT. It is because a unique portable high-resolution mass spectrometer (MULUTM) was used for the first time to break through the intrinsic restriction in MS technique for resolving two major greenhouse gases (GHGs) CO_2 and N_2O by MS technique. This breakthrough paved the way for MS-based *continuous field observation* of *multiple* and *wide variety of soil-related gas*es (*i.e.,* not only GHGs but other gases, for example, O_2, N_2, noble gases, volatile organic carbons (VOCs), and inorganic gases) as MULTLUM has been used for the measurements of various gas species (Jensen et al. 2017, Kawai et al. 2018, Shimma et al. 2013). Recent cavity-ring down spectroscopy (CRDS) instruments offer better measurement precisions for greenhouse gases, as the reviewer pointed out. But the CRDS instruments can measure a limited range of soil gases, including GHGs and single CRDS instrument measures only several gas species.

As the reviewer also pointed out, our MS-based soil gas flux measurement system is in the stage of "proof of concept". In this paper, we intended to show that our MS-based instrument is capable of observing multiple GHGs flux simultaneously in the field, except for low flux conditions at this stage (*i.e.*, except for conditions below minimum detectable fluxes). We are not arguing that our MS-based flux measurement system is equal or superior to CRDS instrument in terms of measurement precision and limit of detection. However, our MS-based technique has the ability to measure multiple greenhouse gas fluxes simultaneously in the field, and, even more, is capable of observing concentrations/fluxes of far more multiple gas species than those single CRDS instrument can measure at a time. To enhance the point, we added the following literature to show wide variety of gas species observed by MULTUM in the text (Jensen et al.,2017, Kawai et al., 2018, Shimma et al., 2013). We understand the reviewers' critique about "not as good as CRDS" for our MS-based instrument at this stage, but as mentioned above, it is beyond the scope of current paper, and a further improvement in our MS-based instrument is future work.

Jensen, K. R., Hondo, T., Sumino, H. and Toyoda, M.: Instrumentation and method development for on-site analysis of helium isotopes, Anal. Chem., 89, 7535–7540, https://doi.org/10.1021/acs.analchem.7b01299, 2017.

Kawai, Y., Hondo, T., Jensen, K. R., Toyoda, M. and Terada, K.: Improved quantitative dynamic range of time-of-flight mass spectrometry by simultaneously waveform-averaging and ion-counting data acquisition, J. Am. Soc. Mass Spectrom., 29, 1403–1407, https://doi.org/10.1007/s13361-018-1967-1, 2018.

Shimma, S., Miki, S., Cody, B. R. and Toyoda, M.: Ultra-High Mass Resolution Miniaturized Time-of-Flight Mass Spectrometer "infiTOF" for Rapid Analysis of Polychlorinated Biphenyls, Comprehensive Analytical Chemistry , 61, 303-323, http://dx.doi.org/10.1016/B978-0-444-62623-3.00013-7, 2013

- We also consider that O_2 *flux* measurement between soil-atmosphere is still challenging even after extensive improvement as pointed out by the reviewer, since measurement precision (currently ± 0.60 vol%), rather than sensitivity, needs to be improved by more than three orders of magnitude. We described various further improvements (*e.g.*, retrofitting a larger vacuum pump to the MULTUM and increase the measurement rate and using a flux chamber with a lower ratio of the height to the bottom area). However, these are going to contribute to enhancement in sensitivity, but not in measurement precision. Although O_2 *flux* measurement is quite challenging, O_2 *concentration* measurement is quite possible with current measurement precession (± 0.60 vol%). In the soil environment, O_2 concentration changes from ~ 0 to 20%, and the current precision of ± 0.60 vol% in O_2 concentration measurement is enough to capture the concentration variabilities.

For example, O_2 concentration is a useful tracer for redox status, which is closely related to sink strength of methane in upland soils as suggested by the reviewer (*e. g.*, Kaiser et al., 2018), and also quite useful to deduce biological status in rice paddy soils (*e. g.*, Lee et al., 2015). Therefore, we removed O_2 from the list of *flux* measurement gases but treated it as a *concentration*-observable gas.

Kaiser, K. E., McGlynn, B. L. and Dore, J. E.: Landscape analysis of soil methane flux across complex terrain. Biogeosciences, 15, 3143-3167, https://doi.org/10.5194/bg-15-3143-2018, 2018.

Lee, H. J., Jeong, S. E., Kim, P. J., Madsen, E. L. and Jeon, C. O.: High resolution depth distribution of Bacteria, Archaea, methanotrophs, and methanogens in the bulk and rhizosphere soils of a flooded rice paddy. Frontiers in microbiology, 6, 639, https://doi.org/10.3389/fmicb.2015.00639, 2015.

- ➢ During another field campaign, we did capture CH_4 fluxes larger the MDF (35.4 µg C m^{-2} h^{-1}), although we did not see any detectable CH_4 flux above the MDF during laboratory experiment and the field campaign presented in this paper. The observed methane flux was negative (soil is sink for atmospheric methane) and tended to increase as soil moisture decreased. We added an example of successful CH_4 flux observation in the field to show that our MS-based instrument is capable of observing CH_4 flux in the farm field under a certain condition (Figure 12).

- ➢ O^+ (*m/z* 15.99) was simultaneously detected along with CH_4^+ (*m/z* 16.03) in a single mass segment (mass range from *m/z* 15.90 to16.05) since they have similar mass. If oxygen is observed as O_2^+ (*m/z* 32.00), another sequence is required, and less measurement time can be allocated for CH_4^+ and N_2O^+ measurement, resulting in lower sensitivity for CH_4^+ and N_2O^+. Due to this reason, we observe oxygen as O^+ in ion counting mode along with CH_4^+.

  More detailed responses are given below.

Materials and Methods

L168: Please clarify what is referred to with the term analytical precision. LOD, RSD? The unit doesn't comply with RSD definition as ion count or peak area. For LOD, parameter k of equation 1 and 1 sigma don't comply.

- ➢ We just forgot to delete the unnecessary sentence. We deleted the sentence of "(L168) The analytical precision (one standard deviation, 1σ) … were obtained at their atmosphere concentrations"

L194: accuracy and precision are different quantities. Why is this "definition" necessary?

- ➢ Since flux is determined by the *rate* of increase in gas concentration in close chamber technique, the definition of minimum detectable flux (MDF) is different from that of ordinary detection limit.

The MDF metric was developed by Christiansen et al. (2015), which was then modified by Nickerson (2016) (eq. 3), and have been used as one of common performance metrics in flux measurements with closed chamber measurements of trace gas flux, in particular, for the flux measurement methods based on continuous gas concentration observation.

$$MDF_i = \left(\frac{1}{t_c} \cdot \frac{A_{a,i}}{\sqrt{n}}\right)\left(\frac{V \cdot P}{S \cdot R \cdot T}\right) \quad (3)$$

The device accuracy ($A_{a,\,i}$) is defined as an instrument's measurement accuracy (Christiansen et al. (2015), Nickerson (2016)). In the flux measurement with CRDS instrument, it is the literature value provided by instrument manufactures, such as Picarro Inc, but no clear definition is publically provided. MDF is a useful metric in the comparison between CRDS and our MS-based instrument, and we employed the MDF for the comparison. In our case, we defined the device accuracy as an ordinary one as described in the paper.

L201: filled instead of spared; what soil mass was used? Was it sieved, or treated in any way, from which depth did you take a sample?

- This soil was taken from 0-10 cm below the soil surface. After the soil sampling, it was sieved to remove roots and stones. We added the sentence in Line 201 as;
  "(Line 201) The soil was taken from 0-10 cm below the soil surface. After the soil sampling, it was sieved to remove roots and stones. The soil was spared in a 60 L plastic container, and the automated flux chamber was placed on the soil.".

L202: I suggest changing to production instead of generation.

- Changed to production instead of generation.

L203: the term "soil gases" is unclear. Do you mean "initiate production or consumption of CO_2, CH_4 and N_2O"?

- Yes. We changed from "generation soil gases" to "initiate production or consumption of CO_2, CH_4 and N_2O" in Line 229.

L300-302: in other words, there is an experiment that supports your notion that fluxes of N_2O, CO_2 and CH_4 can be determined in the field, but you don't show the CH_4 part in the proof-of concept study?!

- We added an example of successful CH_4 flux observation in the same farm field but in a different season to show that our MS-based instrument is capable of observing CH_4 flux in a farm field under a certain condition (Figure 12). We also changed the sentence (L301-302) as;
  "(L301-302) CH_4 flux above the MDF was observed in the same field but during a different field campaign in March 2019. The methane flux tended to increase in the frequency when soil moisture decreased. It was around -100 μg C m^{-2} h^{-1} for CH_4 flux when the soil water content was below 17%."

Results

L266: Please revise terminology. Accuracy is a determined value's deviation from a reference value, precision is reflected by sd.

- ➢ We defined the analytical accuracy in L194 as "We define the analytical accuracy ($A_{a,\,i}$) as the analytical precision (measurement uncertainty) of MULTUM for gas *i* and use the two standard deviations (2σ) obtained from 994 measurements of atmospheric gas as a reference .". We changed the sentence in L266 as;

  "(L266) Their frequency distributions nicely agree with Gaussian distributions (plotted as dashed lines in Fig. 7), and thus their standard deviations are regarded as analytical accuracy ($A_{a,i}$) of the MULTUM–soil chamber system for each gas. The analytical accuracy ($A_{a,i}$) is defined as the analytical precision (measurement uncertainty) of MULTUM for gas *i* and use the two standard deviations (2σ) obtained from 994 measurements of atmospheric gas as a reference, as described in section 3.1."

L270: Why? Please elaborate

L271: Please rephrase the sentence. The meaning is unclear. Are you saying that MDF is not reliable, and, for this reason, you calculated MQF in a reliable way? What is the point in presenting MDF then? Please clarify.

> *(L270-271: Although the MDF represents the minimum detectable flux, it is not a practical measure for reliable quantification of flux. Thereby, we evaluated minimum quantitative flux (MQF) for each gas as quantitatively reliable metric)*

- ➢ We considered MDF is not a useful metric for reliable quantification of flux, which is just a minimum *detectable* value. Thereby, we evaluated the minimum quantitative flux (MQF) in our study.

L274: the standard error of the slope can be calculated. Please refer to Crawley's R book section 10.1.5. Why don't you use the standard error of the slope to calculate the uncertainty of the flux?

> *(L274: The accuracy of MQF depends on the variation of the slope of the regression line. As there is no formula for error/accuracy estimate in such slope (L275) determination, we conducted a simulation study to characterize the MQF considering the measurement error.)*

- ➢ Flux is determined based on the slope (rate of gas concentration change in flux chamber during closed). The data points and associated error bars in Figure 9 reflect the averages and standard deviations of the slopes calculated from 10000 simulated flux measurements for each flux condition. Therefore, the error bars in Figure 9 correspond to those the reviewer suggested (standard error of the slope to calculate the uncertainty of the flux). The reviewer may be talking about applying linear regression to entire simulation results (*i.e.*, 10000 x 8 flux conditions). However, we think it is not a rational approach since adding random error to *true* flux values may not maintain linear relationship between *true* and calculated (simulated) fluxes due to *non-linearly* in linear regression analysis. Therefore, we drew just a 1:1 line to aid readers in grasping their approximate relationship. The description of L274-275 was somewhat unclear and may not easy to understand. We changed it as;

"(L274-275) The MQF is determined from the precision of the slopes (rates of gas concentration changes) in flux measurement relative to *true* slope. However, *true* slopes are hard to know in actual field measurements. We thus conducted a simulation study to characterize the MQF of the current instrument for each gas species.". We also changed the caption in Figure 9 as:

"(Figure 9) Relationship between *true* and simulated fluxes. The error bars in the figures represent two standard deviations of the slopes calculated from 10000 simulated flux measurements for each flux condition. The MQF is defined as a minimum true flux when the true flux is equal to two standard deviations of the slope."

Section 3.4: elements of discussion, but actually speculative, and based on a 5 days campaign after tillage.

➢ When we submitted this paper to AMTD, another reviewer suggested having more discussions/speculations about the observed interesting temporal difference for different gas species. We thus added some speculative discussions, although we also thought that this field study was just for 5 days, and discussion would be quite speculative. We know the current discussion is speculative but quite interesting and worth noting since quite few continuous simultaneous observations of N_2O, CO_2, and CH_4 are available. Therefore, we left the discussion section as it was.

L335-337: It sounds like the authors could apply another data evaluation method, which could turn O_2 measurements feasible. Please clarify why this has not been done.

"(L335-337) *Also, applying waveform averaging mode for the measurement of more abundant O_2^+ instead of current ion counting mode for O^+ should improve the analytical precession of O_2 concentration measurement, and O_2 flux measurement will be feasible.*"

We detected oxygen as O^+ (not as O_2^+) using ion counting mode since O^+ (m/z 15.99) can be simultaneously detected along with CH_4^+ (*m/z* 16.03) in a single mass segment (*m/z* 15.90-16.05). If oxygen is observed as O_2^+ (*m/z* 32.00), another sequence is required, and less measurement time can be allocated for CH_4^+ and N_2O^+ measurement, resulting in lower sensitivity for CH_4^+ and N_2O^+. Due to this reason, we observed oxygen as O^+ in ion counting mode. We added the sentence of why we measured oxygen as O^+ in Line 144 in section 2.1 Simultaneous GHGs and O_2 measurement using MULTUM.

"(Line 144) Oxygen was detected as O^+ (not as O_2^+) using ion counting mode since O^+ (*m/z* 15.99) can be simultaneously detected along with CH_4^+ (*m/z* 16.03). If oxygen is observed as O_2^+ (*m/z* 32.00), another mass segment (around *m/z* 32.00) is required, and less measurement time can be allocated for CH_4^+ and N_2O^+ measurements, resulting in lower sensitivity in CH_4^+ and N_2O^+. Due to this reason, we observe oxygen as O^+ in ion counting mode.

Figure 7: mean and sd of fitted gaussians would be helpful in caption.

➢ We added the mean and standard deviation for each overlaid Gaussians distribution in Figure 7 and changed the caption in Figure 7 as;

“Figure 7. Frequency distributions of measured atmospheric concentrations of N_2O, CH_4, CO_2, and O_2 (994 samples) during laboratory atmospheric air measurement with MULTUM-soil chamber system. For visual comparison, Gaussian distributions, which means (avg) and standard deviations (sdev) were calculated from associated histograms, are plotted as dotted lines.”

Figure 9: 1:1 line would be helpful

The dotted lines in Figure 9 mean 1:1 line. We changed the caption in Figure 9 as;

“Figure 9. Relationship between *true* and simulated fluxes. The error bars in the figures represent two standard deviations of the slopes calculated from 10000 simulated flux measurements for each flux condition. The MQF is defined as a minimum true flux when the true flux is equal to two standard deviations of the slope for each flux condition.”

➢

Figure 11: caption doesn’t explain dashed lines.

- We changed the caption in Figure 11 as;

  “Figure 11. Temporal variations of observed N_2O and CO_2 fluxes at the University Farm of Ehime University during field flux observation in September 2018. The dotted lines represent QMFs. Fluxes below the MDF are masked. The shaded areas represent no data due to measurement interruption by system trouble, and so on.”

- We also added an example of successful CH_4 flux observation in the same farm field but in different season in Figure 12.

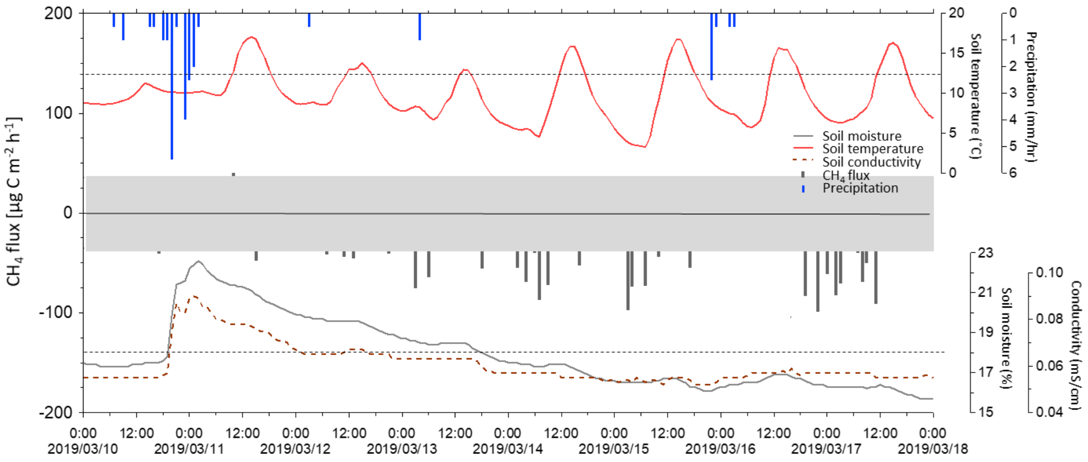

Figure 12. Temporal variations of observed CH_4 flux at the University Farm of Ehime University during field flux observation in March 2019. Fluxes smaller than the MDF (35.4 µg C m^{-2} h^{-1}) are masked. Dotted gray lines represent the MQF (139 µg C m^{-2} h^{-1}).