Brummer et al: Gas chromatography vs quantum cascade laser-based N2O flux measurements using a novel chamber design

The authors have tested the performance of a QCL analyzer connected to a new automated chamber, against a "conventional" GC + automated gas sampling unit system. Data from QCL system were used to observe the non-linearity in the concentration increase during the chamber closure. Based on two short campaigns, the paper gives recommendations how should the measurements, data screening and flux calculations be done. The new chamber design is interesting and the system coupled to QCL seems to be fluently producing nice data. Papers presenting new chamber designs, are always welcome, particularly if they can provide generalizations and recommendations which are useful for other chamber operators. The paper is fluently written, and the observation of different patterns in diurnal cycle is interesting and important. However, there are several deficiencies and pitfalls in the data treatment and the argumentation which need revision. The presentation quality would benefit from separating the results and discussion.

First, the performance of the GC sampling system makes me wonder whether the comparison of two systems is meaningful. Before making any comparisons, the authors should find out the reason for the bad performance of the GC. Secondly, there are several conclusions in the paper which are just qualitative, and as such they are vague and are not supported by the presented data. Third, I share the worry of the first reviewer that most of the results shown here are already well known. For example, it has been reported already in numerous papers that the curvature in concentration increase is higher with longer closure time, and that using the linear calculation instead of non-linear can result in great underestimate in flux rate. Instead of reporting curvature, it would be more useful to quantify what is the limit of curvature after which the authors recommend the use of non-linear fitting method. Also, it would have been interesting to learn more about the advantages and possible problems in the "novel" chamber design. In general, the paper could be more valuable would it provide more quantitative information and recommend some general tests which each chamber operator should run to ensure adequate data quality. It is also a bit questionable if the paper with such a short piece of data (25 + 6 days) is enough the draw firm conclusions.

See more comments below.

MORE DETAILED COMMENTS:

What is the reason for the very bad performance of the GC system? On p5 lines 7-10 it is said that the system was checked against ten samples of ambient air, and only if the CV falls <3%, the data is acceptable. Is this CV limit of 3% really acceptable for a GC system? From Figs. 2a and 7a it seems clear that the GC is not able to resolve concentration increases for fluxes < 20 µgN m-2 h-1. At least to my knowledge, much lower fluxes analyzed with the GC have been reliably reported. I think that a comparison between QCL and GC is not really meaningful if GC is not able to measure these "small" fluxes of N2O. However, Fig. 2a makes me doubt, whether the problem is in the autosampler, and not in the GC detection limit? In some cases the GC can quite perfectly detect a concentration increase of about 12 ppb's similarly to the QCL (second measurement of DOY 339), but during many other closures the data seems arbitrary. What is the reason for that?

There are conclusions in the paper which are not supported by the presented data. For example: p.1 line 30: "new chamber design reduces the disturbance of the soil". There was nothing on that in the results. What are the possible disturbances? How can you detect that those can be omitted by your system? Or: lines 25-26: GC was found to be a useful method to determine N2O fluxes at longer time scale". Where is the data to prove

such a conclusion? There were no budgets calculated. What happens with the low fluxes, how can you reliably determine budget if you cannot detect the flux? Or p.8 line 31 forward: how do you justify the recommendation of removing the first 2 minutes of data? Or "1 to 5 s frequency was sufficient to keep SE on much lower level than in fluxes determined by the GC method" (p. 9). "sufficient" was not defined here. How do you justify the limit of using only the first 10 min of the data? Please give some argument based on the data, not just the feeling that this is good.

Many of the conclusions of the paper follow those observed in previous studies and are already well known. The one exception is the data in Fig. 8 showing the diurnal variation in N2O flux, as this phenomenon is not much studied. In addition to such data providing information on GHG formation processes, the value of the paper could have been in showing how exactly this chamber system works and what are the special and quantified conditions needed to run the system and to screen the data in order to provide reliable flux data.

I strongly recommend to separate the results and discussion parts; presently it is difficult to follow the storyline.

METHODS

P.4 L.3 Why "semi-automatic"?

P.4 L.6-7 A volume of L x W x H does not result in 0.33 m3

Chapter 2.3:

p.5 L. 26: what are the conditions when HMR function cannot be fitted? ; L27, what is Akaike information criterion, please open this a bit, although there is the reference, the reader should get some kind of an idea just by reading the text here.

Equations 1-3 and the text related to them: add units.

RESULTS

P.7 L.7: "low negative k values": care should be taken to express the relations between negative and more negative values. Perhaps more clear to speak about absolute values when comparing these.

P.7. L 10-11 "Near zero fluxes indicate no considerable changes in N2O concentration". Isn't this self-evident without any measurements? Also, what is "considerable change in N2O concentration"? Do you mean significant? If there's no significant increase in concentration, there is no flux, true? Remove or reword the sentence. The whole chapter (Lines 7-14) seem quite self-evident, as the authors hint in the last sentence of the chapter. From Line 15 onwards you say that application of linear model is acceptable in some cases. However, no quantification, i.e. limit below which this is acceptable, is given. I also do not understand how do you draw this conclusion from the results on Lines 7-14.

P.7. L.23: what is meant with dispersion here?

P.7 L. 29-30 "…outside the chamber and inside chamber conditions…" please reword

P.7 L. 30-31 "…coupling of the flux under ambient conditions…" I do not understand this sentence, please reword

P.7 L.28 → What about the impact of the fan speed on curvature? Soil pores may be ventilated also by the fan (see for example Lai et al. 2012, BG).

Chapter 3.2

Might be good to start with your own results, not with the literature review. For clarity, I strongly recommend separating results and discussion.

P.8 L. 15 onwards: Figure 4B indicates that actually the 3-min/lin method produces higher fluxes than the 60 min/exp method in the lower flux regime (below 200 µg N m-2h-1). When taking into account also the higher fluxes (n=6), the relationship changes so that these 6 data points make a very strong impact, as the authors already discuss. Even though this is the case, the discussion here emphasizes continuously how the linear fluxes are smaller than exponential, although the results shown support this observation only for the few high flux points. This makes me to doubt how one can make generalizations about the validity of these two methods. I am missing discussion which tries to find explanation for the higher fluxes with 3-min/lin method. Is it so that the data set should be split, or is it far too small to make generalizations?

Would be also interesting to see, what happens to the SE/RMSE or similar, when apparently low fluxes are calculated with the exponential method. Is there perhaps a risk of higher error /noisy flux data? Would it be possible to find a flux rate below which the linear method is working more reliably than the exponential?

L. 25 I do not understand this sentence

P.8 L.31 onwards, continuing on P.9: Here you give important recommendations, but show no data. Also the reference to Section 3.1 is strange, as I do not find anything about the delayed concentration increase in 3.1. This data should be definitely shown if such recommendations are given. You should justify the removal of the first 2 minutes of data: why exactly 2 minutes?

P.8 L3. "…we also compared HMR-based fluxes **from QCL?** with robust linearly calculated…". How does this vary from that in Fig 7 upper right panel?

A general comment/hint: there are many different comparisons with different analyzers, calculation methods and closure times, in which partly different data sets have been used (low and/or high flux) and it is not easy to follow how do all these small experiments differ from each other or support each other. A separate result section with subsections dedicated to each of these questions might help in that. Now there is lot of text (e.g. Chapter 3.2) and it is difficult to follow the argumentation on logics of the text. Also a clearer division into paragraphs would help the reader. And, as already pointed out, division into results and discussion is needed.

P.9 L. 3-10: In which figure are these shown? Are the slope of 0.97 (lin fluxes are independent) and the HMR fluxes being 22% higher in conflict with each other? How is it possible that now the linear and HMR based fluxes estimated from 60-min data are almost identical (slope=0.97), while earlier you have stated that linear method underestimates the fluxes?

P.9 L 11-18: How were the standard errors calculated?

This section and Figure 5: I think that SE is not an appropriate quantity when estimating the "sufficient" frequency of concentration data. By definition, SE is related to (the root square of) the number of observations. It is therefore evident that if you decrease the frequency and the number of data, you increase the SE. In a case where the random error of the concentration measurement during the chamber closure is constant, the SE will

anyway increase in case the number of observations decreases, whereas the NRMSE, or the error in the flux will not increase. Therefore a better quantity to estimate the error related to the frequency of concentration data is RMSE (or NRMSE).

Your argument "..sampling times between 1 and 5 sec are sufficient to keep SE of fluxes on a much lower level…" is vague. How do you justify that exactly the 1-5 sec limit is sufficient? What means sufficient? How much is "much lower level"? Please quantify and justify this with an appropriate and objective criteria.

P.9 L.12: "..to approx.. one minute,…" isn't it approx. half a minute (25.6 sec)?

Chapter 3.3

P.9 L. 23 To be exact, QCL fluxes are not explained by GC fluxes. They are correlated with GC fluxes.

P.10 L4 What does mean "…no dependency on flux value was observed…" Why should SE depend on flux value? Again, how was SE defined?

P.10 L 14 indicates → indicating

P.10 L 15 forward: "…GC is still useful method to determine soil-atmosphere exchange… at longer time scales.." What is your argument based on? There are no budget calculations in the paper. Averages were reported to be similar, particularly for the small flux regime, but at the same time the fluxes were hardly detected with the GC. Is it correct to say that GC fits for budget studies? How big errors are acceptable in budget studies?

FIG 2 add A) and B) to panels and refer to them in the legend

FIG 4

- Refer to "A, B, C and D" before each legend text parts; "Figure 4. a) Comparison of N2O fluxes… b) Linear regression…"
- The legend text should be shortened. Remove phrases such as "Also shown is…" Figure 4b is showing 60-min fluxes plotted against 3 min fluxes.
- Please remove the text "Riso campaign 2013, Willow…" from the top of each separate panel and add that part of information into the legend text which is not already there.
- Panel B: indicate what are the two lines in the figure? Why are they not direct lines, but show some tiny variation?
- In Fig. 4 and Fig 5, what is the reason to compare 3-min linear and 60-min exponential fluxes? Why not to compare separately the lin vs exp AND 3-min vs 60 min closure times?

FIG 7

- Please use A-D notations, not left/right/upper/lower explanations
- Upper panel: define the lines

FIG 8

- An interesting Figure. What does the error bar denote? Is the diurnal variation significant? Why is the hourly data not shown? Are the points averages from many hours? What was actually the frequency of measurements in both campaigns, I did not find it, but I assumed you measured hourly?