Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2019-396-RC2, 2019

Interactive comment on "Measuring frequently during peak soil N_2O emissions is more important than choosing the time of day to sample" *by* Jordi T. Francis Clar and Robert P. Anex

Anonymous Referee #2

Received and published: 20 November 2019

The authors present observations of N2O emissions from an agricultural research station in Wisconsin over three different seasons at high temporal resolution using a Los Gatos N2O analyzer and automatic chambers in an effort to understand diurnal variability of N2O emissions and whether certain times of day are representative of daily emissions. Emissions were measured from three different plots: in one from April through Oct 2015, one from September 2016 to July 2017, and one from September 2017 through August 2018. They integrate daily fluxes based on a minimum of 11 observations, and then rank the daily fluxes from largest to smallest. They bin these ranked daily observations based on size into 4 or 5 largely overlapping bins: observations amounting to 25% (or possibly 30%) of annual fluxes, observations amounting

Printer-friendly version

Discussion paper

to 50% of annual fluxes (or possibly 25%), and all observations. An additional bin contains the observations ranked from smallest to largest that amount to 50% of annual fluxes. In the datasets containing larger proportions of the observations, there is a clear and relatively well constrained diurnal variability. The large 50% and 25% (or 30%) bins - which represent the largest fluxes ('hot moments') - exhibit high variability and lack clear diurnal variability. The authors conclude that relying on a representative sampling time will not accurately estimate emissions during these high flux dates.

The authors have done a commendable job of collecting high temporal resolution measurements across three separate years – it is a very nice dataset, and I very much appreciate the effort undertaken to obtain it. That said, I have several major concerns.

1. I leave the question of editorial suitability to the editor, but I could use a little more convincing that this manuscript wouldn't be better suited to a journal more focused on agriculture, given the manuscript's narrow focus on questions of sampling representativeness of fertilized fields. With respect to novelty, previous studies have similarly found that N2O fluxes often do not exhibit diurnal variability, particularly when fluxes are low or high (e.g., Laville et al 1999, Van der Weerden et al. 2013), although to say that these particular studies are much shorter in duration would be an understatement, and the long-term dataset included here is certainly of value.

Incidentally, one interpretation of the results would seem to be that PMTs could be the optimal design in instances where sub-daily sampling is not possible: if PMTs accurately reflect daily emissions during the majority of sampling days, it makes sense to use them. If on high emission days there is no diurnal variation, measuring during a PMT is as good a time as any. Perhaps the real message might be that effort should be made to sample hot moments in higher than daily temporal resolution. A question I'm left with is: how does variation in the sampling timing and frequency during hot moments affect annual flux estimates (e.g., comparing a single daily measurement at the PMT to varying sets of additional, sub-daily measurements)? In other words, some quantification of the downside of using a PMT would be very interesting, and germane to the central question of this study.

2. The analyses presented here are focused on the question of sampling representativeness based on an empirical evaluation of when diurnal variation does and does not occur. It does not attempt to explore why and under what conditions diurnal variation breaks down: there's almost no interrogation of process or mechanism. Understanding the degree to which PMTs are representative of daily fluxes is a worthwhile goal, but the presentation here feels thin without substantive investigation of why PMTs are or are not representative. There are many data here that were collected but don't appear to figure at all in the analysis and are not presented as results, in particular depth-resolved soil temperature and moisture observations (Table 1 informs us that some kind of temperature/N2O relationship existed, but no details on that relationship are presented). Seasonal variation in the duration and timing in PMTs is referred to in passing but not actually presented. There is no consideration of soil C or N. There are a lot of questions I was curious about while reading the manuscript. For example, does diurnal variation in high fluxes depend at all on whether the pulses are related to fertilizer applications, freeze-thaw events, precipitation events, or some other cause? Do organic and inorganic fertilizer applications affect diurnal variation differently (through effects on soil O2 and organic C availability to denitrifiers)? How does (seasonal) variation in the range of diurnal temperature variation affect diurnal variation during high flux events? Even with the three-year data set there may not be enough replication of events to answer some of these questions statistically, and the organic v inorganic question can't be analyzed statistically with the current experimental design, but I would think some quantitatively-informed discussion would be possible, and could be a good way to take advantage of this very nice dataset.

3. I have a number of concerns with the binning approach used.

First, the description of the normalized cumulative daily fluxes in line 274 and following is not very clear and could be done much more simply: I'm not sure it is necessary to define a new concept here (It's also not necessary to normalize, especially since the normalization is not actually carried through in the analysis: the results (Figure 1) present raw fluxes, not normalized fluxes). If I understand this paragraph correctly, 1) you ranked daily fluxes in reverse order based on magnitude, from largest to smallest 2) you binned these ranked fluxes based on their contribution to total annual emissions. Bin 1 (75% High Cumulative Contribution (HCC)), Bin 2 (50% HCC) and Bin 3 (30% HCC) contain your ranked daily fluxes cumulatively representing 75%, 50%, and 30% of annual emissions, respectively. I initially thought Bin 4 (50% Low Cumulative Contribution (LCC)) was the collection of daily fluxes not included in Bin 2 (50% HCC), but if 50% LCC represents 85% of measurements, that would seem to suggest that instead it might actually be 25% LCC (i.e., the bin selected from daily fluxes ranked in increasing order to sum to 25% of the total cumulative flux), since 75% HCC contains 15% of all daily fluxes (in which case, Bin 4 is actually the collection of daily fluxes not included in Bin 1). It's also non-intuitive that a large number of daily fluxes are being described both as High Cumulative Contribution and Low Cumulative Contribution (the overlap in the 75% HCC bin and a possibly hypothetical 50% LCC bin). And the "50% LCC" term is not included in the Results, Discussion, or figures, though the results presented in Figure 2 appear to represent 50% LCC (though again, it may actually be 25% LCC).

Figure 1 also suggests that there are further factual errors in the binning description in the methods. Figure 1 suggests that there are breakpoints at 75%, 50%, and 25% of the annual flux, creating 4 bins of daily fluxes, though a given daily flux may be present in more than one bin. This suggests that there is no 30% HCC bin, though this is not strictly clear. The results section refers sometimes to a 25% HCC bin, and sometimes to a 30% HCC bin. The caption for Figure 1 first refers to panel D as being based on 25% of the total flux, and later as being based on 30% of the total flux. There is no mention of the 25% HCC bin in the methods.

Your bins often overlap with one-another, which is a bit unusual, and the bins aren't structured in a systematic or symmetrical fashion (30% HCC overlaps with the other two HCC bins, 50% LCC and 50% HCC overlap with 75% HCC). There are also statistical issues with overlapping bins that need to be kept in mind: because the same observation may be in multiple bins, the bins are not independent, and so comparisons between bins violate any assumptions of independence. In the end, if both overlapping and non-overlapping bins are used, it might be helpful to provide a rationale for that structure as opposed to the alternative of non-overlapping bins. Non-overlapping bins would have the added benefit of being able to explicitly evaluate how variation in flux magnitude affects diurnal patterns, something I would argue the current analytical design is not capable of doing because of the overlap in bins. In addition, the current analysis does not investigate the possibility that diurnal variability is absent when emissions are low and if so, why. This is a question that may not be important for estimating annual fluxes, but is relevant to our scientific understanding of diurnal variation in N2O fluxes.

There's also an important statistical issue with using bins having unequal numbers of observations: the standard error is directly proportionate to the sample size. You use comparisons of the error in Beta to argue that there is no diurnal variation during hot moments, but because of the large differences in sample size, it's not surprising that the estimates of Beta in panel a of figure 1 –which is based on all of the measurements (n= 2,017 days)–or in Figure 2, with 85% of all measurements (n > 1,700 days), have lower uncertainty than the estimates of Beta in Figure 1 panel c (n= 55 days, or 6%) or Figure 1 panel d (n = 27 days, or 3%). Bins of equal size would address this issue. If there is a reason to keep unequal bin sizes, it would seem to be important to at least show that bins containing 3% and 6% of measurements centered around the median exhibit clear diurnal variability, and probably good to evaluate the lowest 3% and 6% of measurements given the possibility that diurnal variability patterns break down at both low and high fluxes. It's essential that this analysis be robust, since at the moment it's the central finding being promoted in the manuscript.

Specific comments:

Introduction: The Introduction provides a review of articles that have evaluated diurnal variability and whether a PMT is reliably observed in a laundry-list format. It is not strictly necessary, but the authors might consider whether it would be possible to provide a greater synthesis of the main results of those studies, and present details in a table. This would have the added benefit of freeing up word count to provide more mechanistic context in the introduction, which is needed but currently lacking: specifically, mechanism behind diurnal variation in N2O emissions from agricultural soils, and what causes that diurnal variability to break down during high emission periods. In addition, “Hot moments” is used multiple times in the abstract, but is missing from the introduction. If "hot moments" is going to be used as a key framing device in the abstract, it also needs to be introduced and contextualized in the introduction. Otherwise, it should be removed from the abstract (and keep in mind that "hot moments" is jargon, though very evocative jargon!).

Line 131: It’s not clear how positioning sites within 2.25 kilometers of one another addresses possible effects of manure amendments. It would be helpful to have some clarification, including description of what those possible effects would be.

Line 143: if there’s a reference for results from this study, please include.

Line 145-6: “In each campaign, both treatments campaigns dairy slurry was applied” needs to be re-written, perhaps delete “both treatments campaigns”

Line 193: change “valve assembles” to “valve assemblies”

Line 214: Change “Flus” to “Flux”

Line 240: It might be helpful to have a rationale for why these fluxes were removed, rather than included as a net zero flux. If any negative fluxes were excluded, a rationale for that would be needed.

Line 261: Open parenthesis is missing

Line 270: this paragraph should be placed after the normalized cumulative daily flux has already been defined, since it is the independent variable in the analyses.

Line 275: insert “of” before “the annual”.

Line 276: how does the “cumulative daily flux” differ from the “daily flux” defined at line 257? If it does not, please use the same term for both; I would suggest sticking with ’daily flux.’

Line 276-277: there’s a problem with the sentence construction (“data subsets using to the normalized. . . “).

Line 290 and 291: insert commas in the thousands place for 1,093 and 2,017

Line 293: The abbreviation “HCC” has already been introduced

Line 293 and 295: the use of “measurements” here is a little ambiguous. It might be helpful to clarify whether a “measurement” refers to an individual day, or to an individual (sub-daily) flux. Also not sure reporting a mean of 912 provides useful information to the reader if it’s a mean of 5 values ranging from 27 to either 2,017 or over 20,000. Reporting the number of measurements in each bin makes more sense.

Lines 313-316: as noted in my major comment 3 above, the possibility that these results may simply be caused by the huge difference in sample size needs to be resolved.

Line 320: I don’t think you’ve quite shown this yet, because 1) of the issue of having different sample sizes for different flux intensities, and 2) the design of your analysis is not quite an investigation of flux intensity - i.e., different flux intensities are not compared (with the possible exception of 50% HCC and 50% LCC). Also, delete “to”.

Line 321 and rest of paragraph: I might think of another way of contextualizing what you call “low” here, since you aren’t analyzing fluxes that are low in the context of, for example, your mean flux, but relative to the highest $\sim$10% of fluxes.

Line 330 and following: it would be nice to actually see the results you mention in these lines, as well as some analysis and interpretation of them.

Line 336: Emissions of 723 g N2O-N ha-1 day-1 seem to me to be quite high for a

Printer-friendly version

Discussion paper

‘low’ classification, though you do indeed have some very high individual fluxes. But it seems potentially problematic. For example that Laville et al. 1999 - cited in the manuscript as an example of high fluxes with no diurnal variation–observed maximum hourly fluxes of 700 ng N2O-N m-2 s-1, which extrapolates to a daily flux of roughly 600 g N2O-N ha-1 day-1, falling into the “low flux” category of this manuscript.

Line 364: again, not sure your analysis allows you to say this. You could perhaps just say “diurnal variability of N2O emissions is not present during the largest emission events” under the current analysis.

Figure 1: “N-N2O” is used here, but in the text “N2O-N” is used. Units in panel D (1000, 2000) are mislabeled as 000. Worth a quick double/triple check of concentration measurements and flux calculations to ensure the units are correct since the fluxes are on the high end (many measurements at 1.5 to 3 kg N2O-N ha-1 day-1 during pulses). Delete "and" from "These account for and 100%"

Figure 2: These results could alternatively be included, along with daily flux measurements, as a separate panel in Figure 1.

References: It would be helpful to readers if all references are either indented at the first line, or if a carriage return is placed between references. Scheer et al 2012 Plant Soil would be a good addition, as it includes both high-frequency measurements and evaluation of diurnal variability. DOI 10.1007/s11104-012-1197-4

Line 469: wrong location for Cosentino et al. 2012.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-396, 2019.