Response to Anonymous Referee #1 (Referee comments are shown in *Italics*)

We thank the referee for their valuable comments that will help improve the manuscript. Before responding to the referee's individual comments we want to clarify that the aim of this paper is not to elucidate the mechanisms responsible for the diurnal variability of soil N_2O emissions, but to provide other researchers with knowledge that will increase the accuracy of their annual emissions estimates and inform the optimization of their sampling protocols. The design of soil trace gas emission monitoring experiments is often based on studies of diurnal cycling of emissions made with very limited data that did not capture seasonal or annual variability, and in which 'hot moment' emissions are absent or overlooked. Our goal in this paper is to characterize the temporal patterns of emissions, rather than explain mechanistic cause of those patterns.

Below we address the referee's concerns individually. Corresponding changes will be made to improve the manuscript.

General comments (1)

Authors state several times that this is the first study of N2O flux temporal variability that includes several hot moments. Which isn't really true, is it? There is Luo et al. 2012 (Decadal variability of soil CO2, NO, N2O, and CH4 fluxes at the Hogwald Forest, Germany), and a lot of other studies from that same study site that show temporal variations in N2O. There are also numerous papers from Australia (e.g. Barton et al. 2007 Nitrous oxide emissions from a cropped soil in a semi-arid climate) – although, to be fair, the Barton paper did not really experience what could be called "hot moments". And there is also the Machado et al. 2019 paper (as cited in the current manuscript) that also measured temporal variability during periods that included "hot moments".

Response (1)

We claim that our publications is the first study of N_2O flux temporal variability that includes multiple difficult-to-measure peak emission events (i.e., "hot moments") and estimates the relative contribution of hot moments to cumulative emissions (lines 12 to 15). It is true that occasional larger emissions periods were observed in the publications cited by the reviewer (Barton et al., 2007; Luo et al., 2012; and Machado et al., 2019). The 'high' emissions reported in these publications, however, are approximately two orders of magnitude smaller than those observed in our experiments. The lack of continuous measurement in these studies makes it impossible to determine the importance of these larger flux periods relative to cumulative emissions.

Luo et al. (2012) observed multiple hot moments and estimated their relative contribution to cumulative emissions. This publication states that pulse events that occurred during soil thaw in 1996 and 2006, accounted for 88% and 87% of the total annual emissions, respectively. Luo, et al. (2012) did not attempt to characterize diurnal variability, however, but rather characterized the variability of N_2O emissions at the seasonal, annual and decadal scale. We have not found a

publication in which the temporal variability of N_2O emissions from the Höglwald Forest, Germany have been studied at a sub-daily frequency or with the purpose of providing a sampling time that best represents the mean daily emissions.

General comments (2):

Also, I don't agree that constraining sampling to particular times of the day provide little benefit. Generally, researchers will be sampling during regular working hours (i.e. between 9 am and 5 pm), during those hours, sampling before 10 am will underestimate fluxes (for 50% of annual flux – i.e. see figure 1c), while sampling after 2 pm will overestimate fluxes. So your preferred sampling time should be between 10:30 to 13:30. No? Even figure 1d makes a strong case for not sampling in the afternoon (between 15:00 and 18:00), because it will overestimate fluxes. That being said, I do agree with you that frequency of sampling during hot moments is more important than what time the sampling took place during those hot moments.

So, realistically, I don't think that your conclusions are actually substantiated by your own data. I think that your data still backs up previous research that suggests avoiding the afternoons when sampling for determination of N2O fluxes. Not the most novel conclusion, but I think it is still worthwhile.

Response (2)

We do not dispute the conclusions of previous research about the Preferred Measuring Time (PMT) when a diurnal cycle is observed, rather we conclude that it is *much* more important to sample frequently during peak emissions than to sample at a specific time of day. The recommendations of a PMT in previous research, based on limited data engender false confidence that sampling daily at a particular time is sufficient to yield good estimates of daily and cumulative emissions. We emphasize that such confidence is misplaced.

Previous research is generally based on short experiments that did not include periods of significantly high emission. When a diurnal cycle of emissions **is** observed, this previous research concludes that there is a PMT in the late morning. Our results do not refute this conclusion, but more importantly, they show that these periods when there is a PMT are relatively unimportant in terms of estimating cumulative emissions. A much larger fraction of the cumulative emissions occur during high emissions periods when there is no diurnal cycle and no PMT, and during these periods sampling frequently is essential.

High emissions periods (25% HCC), represent only 6% of the total observations but 25% of the cumulative emissions. On average across the three years, the beta coefficient for each hour during the high emissions periods is computed from data gathered on just 27 days (Lines 291-295). One day's worth of measured emissions during peak emissions is 0.93% (i.e., 25%/27) of the cumulative emissions. On the other hand, low emissions represent 50% of the cumulative emissions, and on average, emissions data from 912 days are used to compute the beta coefficients (Lines 291-295). One day's worth of measurements during low emissions (LCC 50%) is only 0.056% (50% / 912) of cumulative emissions. Thus, getting one accurate measurement

during the high emissions periods is more important than measuring 20 times during low emissions periods (the ratio between 0.93% and 0.056% is 16.5).

We conclude that sampling frequently during peak emissions is more important than sampling less frequently (e.g., daily) at a specific time of day.

General comments (3):

Also, I would like to see a bit more in the discussion about why diurnal patterns are less relevant during periods of high emissions. There is very little on what mechanisms of processes actually drive this. In Figure 1c and 1d, there is much higher variability between 0 and 8h (compared with the rest of the day). Any ideas why this might be? Is there less production? Or is it related to climatic conditions at night?

Response (3)

We would like to be able to provide more insight into the mechanisms of the processes driving the observed patterns of emissions. We invested quite a bit of time looking at weather data, soil and moisture data, and emissions, but were unable to find consistent and reliable explanations for observed emissions in our data. Previous research has linked environmental variables (i.e. soil/air temperature, water filled pore space) to N_2O production but also concluded that there are site- and event-specific factors that drive the biological processes (i.e. depth of N_2O production, soil characteristics, available carbon) and that we did not measure.

We did not explore why there is much higher variability during the late night and early morning (hours 0 to 8) than during the rest of the day at 50 and 25% HCC fluxes. According to the right plots in figure 1, panels C and D, fluxes measured from 0 to 8 are usually smaller than the mean daily flux, but the high variability of the fluxes indicates that this is not always the case. We believe climatic conditions to be influential; nonetheless, other variables that we did not measure are needed to draw meaningful conclusions in this regard (see Shurpali, et al. 2016 and Thies et al., 2019).

What we are able to do with our extensive data set is "to evaluate the use of PMTs as a strategy to improve the accuracy of soil N_2O flux estimates or reduce the necessary frequency of flux measurements in highly fertilized crop systems", lines 120 and 121. We too would like to be able to say more about the mechanisms underlying the observed emissions, and although we cannot, we feel that what we have been able to show about the relative importance of sampling daily at a PMT compared to sampling frequently during high emissions periods is important.

General comments (4):

Finally, try to avoid paragraphs that consist of 1 sentence (e.g. lines 159-161).

Response (4)

Thank you. We agree. We will make appropriate changes and try to do better in the future.

Responses to specific comments

Specific comment	Response
Lines 36-38. The topic of sampling frequency and quantifying what it means in terms of uncertainty for your cumulative estimate has been covered well in the paper by Barton et al. 2015 (see your own citation list), and should probably be cited here.	Accepted
Line 46: I think it would be worthwhile to cite your Table 1 here, because the table does a good job of summarizing some of the literature on timing and existence of diurnal patterns in N2O fluxes. Although, it doesn't seem like there is that much disagreement. Pretty much all of the studies summarized in Table 1 (11 of the 12 that suggest a PMT) suggest avoiding sampling during the afternoon. That is pretty good consensus in my opinion.	The table will be cited. Avoid afternoon sampling is accepted but most likely not to be mentioned because it could create confusion (i.e. avoid afternoon sampling = to accurate cumulative emissions estimates)
Line 214: "Flux" not "Flus".	Accepted
Line 239: wouldn't it make more sense to report your MDF as a flux per hour? Rather than per day? You are measuring flux rates based on 20 min deployment times and are calculating your daily fluxes by integrating the individual flux measurements for that particular day. So it is possible to have some fluxes below the MDF and others above the MDF on the same day. Line 239: wouldn't it make more sense to report your MDF as a flux per hour? Rather than per day? You are measuring flux rates based on 20 min deployment times and are calculating your daily fluxes by integrating the individual flux measurements for that particular day. So it is possible to have some fluxes below the MDF and others above the MDF on the same day.	We reported all fluxes in the same units to facilitate comparisons within the publication. Fluxes below the MDF were removed (line 240), deleted. It is possible to have fluxes below and above the MDF in the same day. Nonetheless the MDF is very small and the chance of fluxes below this size is very small. We believe that most of the fluxes below the MDF are the result of a chamber not closing during the measuring time. We have observed that a large number of fluxes below the MDF occurred during rain periods when the chambers remain open and others occurred when the chamber did not close due to a mechanical fault. Differentiating between mechanical fault and fluxes actually below the MDF is not always possible.

Line 266: the “+0” is unnecessary.	Accepted
Line 275: “percentage of the annual”.	Accepted
Line 289: what happened during 2016-17? Having to throw out 48% of the fluxes is not good (or were these removed because they were below MDF?). I think it might be better to differentiate when fluxes were thrown out because of bad data and when they were thrown out because they were below MDF.	Differentiating between a mechanical fault and a flux below the MDF is not always possible. 2016-17 was our first sampling season through the Wisconsin winter. Multiple mechanical failures reduced the availability of high temporal resolution data during this period. The large number of mechanical failures resulted in the design and construction of new automatic soil chambers. Most of the mechanical problems occurred after an unusual event when the experimental field flooded in late January followed immediately by freezing temperatures that coated the equipment and soil surface with a thick layer of ice.
Line 318: Are you sure that this is the first? Doesn’t the Machado et al. 2019 paper measure diurnal variability in a fertilized agronomic system with hot moments?	Responded previously.