Below please find responses to reviewer comments and author responses (bold, italic).

Referee #1

The work of Savage and colleagues presents data from two experiments using an automated static chamber system combined with laser spectroscopy; providing a highly precise measurement system, together with high temporal resolution. Results reinforce the need of high temporal resolution measurements, due to short-lasting pulses, and show diel patterns of the fluxes, both issues being relevant for extrapolation and accounting. Furthermore, the precision of the system allows the confident detection of soil N_2O uptake processes, which usually fall within the detection limit of most of the measurement systems. Finally, the paper makes a comparison between this automatic system and a "common" manual chamber system.

Although the paper is quite descriptive, the topic is of high interest for the journal and provides some useful insights into temporal and spatial dynamics of soil focus on N_2O and CH_4) fluxes. Unfortunately, the main conclusion I may extract out of this work is that both many replicates and high temporal frequency are needed for reliable estimates of soil GHG emissions.

Overall quality of the paper is good: it is well structured and written, methods are well explained and results and discussion are sound. However, I see a couple of issues requiring further explanations prior to publication. This would avoid misinterpretation of results and make the paper even more useful for people aiming at precise and accurate soil GHG fluxes estimations.

Briefly, I think your minimum detectable flux is indeed larger than what you present; recommendations upon diel patterns may be counterproductive and some issues when comparing manual and automatic chambers may be missed in the text. Furthermore, some details of minor importance are commented.

Minimum detectable flux and comparison between manual and automated systems: The minimum detectable fluxes (MDF) you present are astonishing low! So low, that some clarifications are needed about the procedure to calculate the MDF. I am confident about the reliability of your system, which has a lot of pros and little cons, but I doubt the minimum detectable flux is as low as what you present. Just a rough calculation: the MDF of 0.01 ug N_2O-N m-2 h-1 corresponds to an increase of approx. 1 ppt N_2O min-1 in the headspace, beyond what the QCLAS is capable of.

First, I am not sure whether the CV can be determined by collecting ambient air (P18287, L5), given the intrinsic variation in the mixing ratios of CH_4 and N_2O of such ambient air before the headspace is closed. This is probably the reason why you end up with different MDF in the two locations, but this does not represent variations in the detection limit, or in the sensitivity of your system, but the stability of the composition of the air. According to your text, you follow the procedure of Parkin et al (2012) for calculating detection limits. However, this work shows tables to calculate MDF from three and four time points measurements, only, and not 240 points-measurements. Have you run your own Monte Carlo simulations to develop your own MDF of the system? If this is the case, what was the influence of the number of points taken on the MDF? This is important since the magnitude of the number of points taken is different according to Parkin (3 to 4) and this paper (240).

However, this is likely the theoretical precision of your measurement instrument, but not of your experimental setup (this is mentioned in the paper from Parkin et al.). You should address the fact that you change the environmental conditions in the chamber headspace (10 minutes though, but they change), you may have sources of uncertainty in the system, due to pressure variations when switching valves, etc.

Calculation of MDL:

    We first presented the instrument precision in the manuscript by measuring the mean and standard deviation of high precision NOAA gas standards not run in-line with the automated chamber system. The QCLAS performed with high precision and very low standard deviation about two different gas samples at a sample frequency of 1 Hz.

    We followed the method of Parkin et al. 2012 to calculated MDL as it was a suitable method for the manually sampled fluxes (which had 4 samples of headspace concentration over the sample period) and felt that this would provide a comparison for the

Figure 1: % of confidence intervals that do not include zero –upper two panels blue line North Dakota, lower two panels red line Howland forest using the Verchot et al. (1999) method. Where the solid black horizontal line indicating 67% intersects the colored lines provides the estimate of the MDL. For comparison, the shaded area represents MDL calculated by Parkin method.

automated system as well since it would include instrument precision (QCLAS or GC) as well as errors associated with the automated chamber system (ie chambers, tubing, flows) and the manual chamber system (ie vials). Ambient for the automated system was measured from an unused chamber top in the field which was left in the up position. Ambient for the manual system was collected in vials at the site and run in the same manner as the flux gases samples. For the manual technique we used the Parkin method for 4 sample points and since their method does not include a method for greater than 4 sampling points and we wanted to compare MDL calculations, we used that same method for the automated system technique. This should yield a conservative estimate of the MDL for the automated system with many
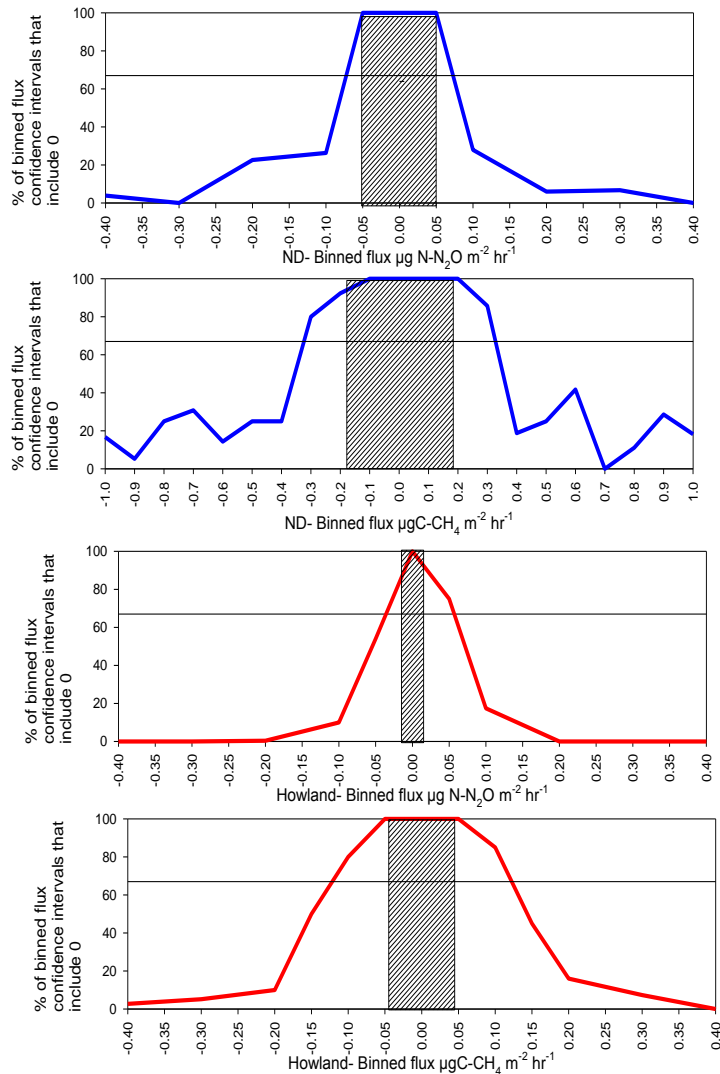
more sampling points, because as the number of sampling points increases, the MDL should go down.

We have looked at calculating the MDL for the automated system using two other methods. The first method was presented in Verchot et al. 1999(GBC 13:31-46), and calculates the 95% confidence intervals around the slope of the change in headspace and then bins fluxes and defines the MDL as the flux bin interval at which >67% of the calculated fluxes have confidence intervals that do not include zero. We binned our fluxes at 0.05 µg N m^{-2} hr^{-1} intervals and calculated the % of confidence intervals within each bin that include zero (Figure 1). Similar to Verchot, we used 67% as the cutoff for determining MDL. The MDL's calculated using the Verchot method are greater than those of the Parkin method, but still very low.

Second, we calculated the MDL using the std of the NOAA calibration gas divided by time over which chambers are usually sampled, (e.g., 0.04 ppb/240sec)to calculate the minimum slope for the change in headspace concentration that must occur to be sure that the change in concentration is detectable, and then the flux is calculated for that slope as for a regular chamber flux.. This produced MDLs for CH_4 of similar range to the Parkin and Verchot method and slightly higher for N_2O than those methods (Table 1).

Figure 2 shows an example of a chamber flux from the Howland forest site with a flux at the detection limit of -0.11 µg N m^{-2} hr^{-1} (95% confidence ±0.03 µg N m^{-2} hr^{-1}, R^2 = 0.10, p<0.0001), which is equal to the MDL calculated from the NOAA standards, based on a change in headspace concentration of approximately 0.04 ppb during 240 seconds (the stdev for N_2O from NOAA standards was also 0.04ppb).

Site	Gas	Parkin method MDL	Verchot method MDL	NOAA-Cal Gas using stdev
Howland	N_2O	±0.01	±0.05	
Howland	CH_4	±0.03	±0.12	
North Dakota	N_2O	±0.05	±0.07	±0.10
North Dakota	CH_4	±0.18	±0.32	±0.25

Table 1: MDL calculations using three methods Units µg C m^{-2} hr^{-1}, µg N m^{-2} hr^{-1}

One of our primary conclusions is that the low MDL from this QCLAS/automated chamber system is that we have confidence in the measured low negative fluxes of N_2O primarily measured at the Howland forest swamp site, most of which fall between -0.1 and -1.5 $\mu N\ m^{-2}\ hr^{-1}$. This is true whether we use the Parkin or Verchot method of calculating MDL
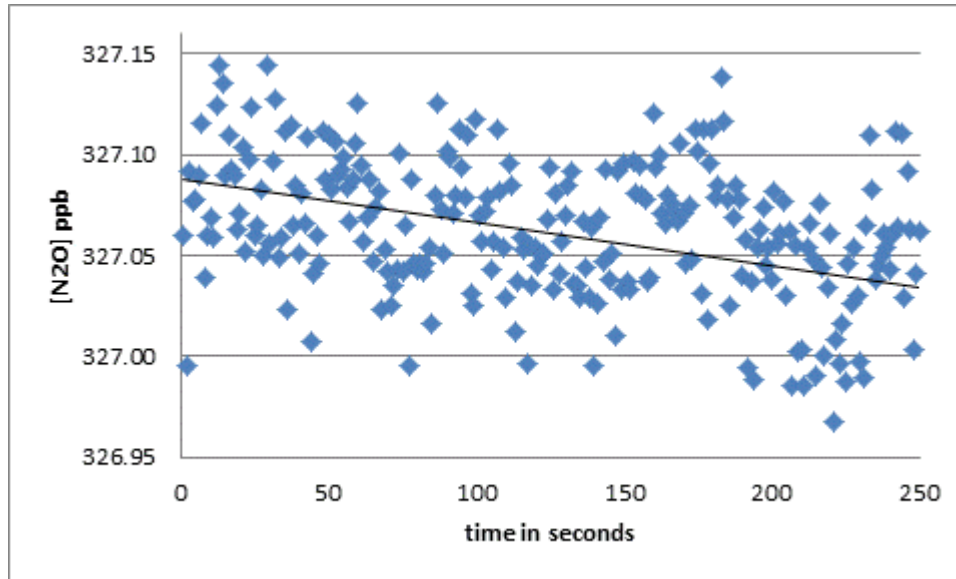

Figure 2: Example of a chamber change in headspace [N_2O] at MDL for the Howland Forest wetland site. Flux is -0.11 µg N m^{-2} hr^{-1} (95% confidence ±0.03 µg N m^{-2} hr^{-1}, R^2 = 0.10)

Diel patterns:

I would be extremely careful with your conclusions. You say the time of day you manually measure (9-12:00) is representative of the mean daily flux. There are a couple of issues here: first, the optimum time is not the same for the three molecules, so that you need a compromise there; second, this morning period is the one with the highest temporal variations in the fluxes. A time gap of 1-2 hours may mean 50 % of difference in the measured fluxes (the 50% is a rough estimate from what I see in Fig 7a), and non-synchronized measurements may be biased. When dealing with several plots and limited man power, measurements are usually performed sequentially, so that a group of chambers is sampled within a period of 60-90 minutes (including closing time + preparation) and when this first group is sampled, the next one is sampled in a similar way, so that there is a 60-90 minutes time interval between both samplings. Thus, another solution may be to take the gas samples very early in the morning, or in the afternoon and then apply a correction factor for the under/overestimation. Unfortunately that is usual not that straightforward when working under field conditions.

The reviewer is correct that in a study with multiple plots that each require an hour or more to sample, it is impossible to standardize the exact timing of sampling. For the purposes of this study in North Dakota, howerver, there was only one plot and all manual chambers were sampled over the same time interval at each sample date. The diel pattern for each GHG was developed from the automated system and took the average GHG flux per hour over the entire crop rotation. We estimated from these diel patterns the time of day representative of the mean daily flux for each GHG. This time period coincided with the time period that the manual chambers were sampled. This result is simply an average time each day representative of the mean daily flux for each gas over the season and may not be applicable to individual

days, which the reviewer correctly notes may range a few hours on either side of this mean on any given day. We do not mean to suggest that sampling at a precise time of day will always provide a good estimate of that day's mean, but rather that, by sampling consistently in the late morning, the long-term average will generally be representative of the long-term daily average flux. We have clarified this in the revise manuscript.

Comparison manual and automated chamber system: In your conclusions, I don't think you can say that "manual sampling : : : captured transient responses of GHG's to precipitation events. This is not what I see in figure 4.

The relatively frequent manual sampling strategy of 3 times per week adopted for this site was able to capture days in which higher rates of GHG fluxes occurred as a result of precipitation events. As is shown by the automated data, the length of response of GHG fluxes to precipitation events is often less than a few days, so manual sampling 3 times per week generally reveals the ups and downs of the flux pattern. The automated system's high temporal frequency more clearly defines the rise and decay characteristics of GHG response to precipitation wet up and dry down of soils. These results indicate that for this particular soil environment in North Dakota, the currently adopted manual sampling strategy is sufficient to capture general responses of GHG's to precipitation events. However if the goal is to quantify the summed hourly effect of a precipitation event pulse and subsequent decay, then the automated system is more appropriate.

Your estimations with manual and automated chambers differ by 300% (26 vs. 9 ug m-2 h-1) and this is something needing more discussion!

The large difference in the seasonal estimate between manual and automated N_2O flux is the result of the high spatial variation of "hot spots" and temporal variation of "hot moments" and how they are captured and characterized differently between the manual 3 times per week sampling strategy and the automated hourly sampling strategy. The effects of capturing a "hot spot" or "hot moment" in the manual sampling can have a greater effect when interpolating between data points over several days' time. In contrast, the automated hourly sampling measures each hour of the day and sums hourly estimates for a daily flux, with only occasional interpolation between points when there are some hours with missing data. Therefore the influence of any one "hot moment" on the overall seasonal estimate is not as great for automated measurements compared to manual measurements. This is addressed in section 3.6 and 3.7 of the manuscript

Linking this issue with the MDF one: How can you explain that with the automated system, as much as 9-12 % of your fluxes were below the detection limit (0.01 ug N m- 2 h-1) but, with the manual chambers, only 5% where below the detection limit, which was 70 times higher! To me, this demonstrates there is a source of variation between both methods, potentially leading to a bias when comparing them and this is source of uncertainty is not explained by the calculation of the detection limit.

These estimates for Howland forest and North Dakota were mistakes in our submitted manuscript. The MDL using the Parkin method was accidently counted higher than determined. Using the correct MDL there were 1% below MDL determined by the Parkin method for Howland and <1% for North Dakota. However, if we now apply the somewhat higher MDL estimate from the Verchot et al. method or the method using the standard

deviation of measured standards, then a higher percentage of nonsignificant fluxes would be correct. We will clarify this in the revision. Regardless of which MDL is chosen, however, most of the negative measured N2O fluxes are significantly above the MDL.

Specific comments:
Flushing: You're flushing your headspace chambers with a total flow of 8 liters during one deployment cycle, which corresponds to 2 times the volume of the headspace. Don't you have pressurization problems? Have you checked this issue?
Our chamber tops have a vent to alleviate pressure due to flow between the chamber and instrumentation and we keep the flow rate to 1L/min. We have tested how the opening and closing of the chamber top and the flow rate affect pressure inside the chamber (Davidson et al 2002, Ag.For. Met., 113,21-37) and it is minimal after the chamber top has closed for this system.

18284, L24: Did the instrument feature a water broadening correction factor, or only pure dilution was taken into account? Was this correction factor able to eliminate interferences due to changes in water content, so that you don't have a source of uncertainty?
Water broadening correction factor was used to correct gas concentration.

P18278; You focus on soil moisture at the beginning of your abstract, but this is not an issue of major importance in this paper
We address the response of GHG's to precipitation events in section 3.5 and this is an important result.

P18284: Is your laser really cooled to 32_C? Is it a typo, or maybe your thermocube is set at 32_C, but not the laser?
This is correct, the laser is cooled to 32°C which is its operating temperature.

P18285, L26. What do you exactly mean when you write that the instrument is trained on interpreting the N2O gas peak? To my understanding, the instrument makes use of the HITRAN database to simulate size and shape of absorption lines, and this information is used for performing the fitting (with help of the tuning parameters) and estimating the mixing ratio of the target molecule. Therefore, I don't get this "training" issue. I don't know which lines are being measured with this specific instrument. Maybe the strength of the CH4 lines is lower? Furthermore, this company usually provides precision according to different sampling frequencies.
This instrument does make use of the HITRAN database to simulate the size and shape of the absorption lines, but there are small manual adjustments to the tuning rate parameters that can be made to improve the model fit to gases. "Trained" was not the correct term; "tuned" is more appropriate to use and has been adjusted in the text.

P18287, L9: "0.18" instead of "0.180"
Changed

P18280, L25. QCLAS corresponds to "Quantum Cascade Laser Absorption Spectrometer". Fig 1. Replace "cool laser" by "laser cooler" or similar. You say only one chamber is closed at a time, but the graph shows chambers 1 and 2 closed
Figure 1 has been updated.

Fig3, 4 and 5. Please homogenize the X Date axis, the format is not the same across the figures (separator, inclusion of year, etc).
Dates homogenized for fig 3,4,5

Fig 4. Do not use abbreviations for the experimental site.
Changed abbreviation in Figure 4 caption

Fig 5. Specify where the data come from.
Added year to sampling date and data are from ND- added this to figure caption

Fig 6d. Scale should be adapted to the magnitude of the values showed, as in the rest of the sub-figures.
Scale has been changed

Fig 6e. I don't see the line for the mean of the "automated concurrent" seasonal N2O flux.
It is there near bottom

I see there was some discussion on the units of the fluxes. I recommend to keep them as they currently are.

Anonymous Referee #2
This manuscript details a new combination of automated, high temporal resolution sampling technology with laser spectroscopy, to attempt to accurately and precisely quantify three major anthropogenic GHGs, with particular emphasis on N2O and CH4, the predominant gases associated with agricultural systems. The authors 'road test' this technology in two contrasting environments to provide a range of sampling and environmental conditions, and compare their 'high-tech' system with a 'conventional' manual sampling approach in an agricultural crop system to investigate the pros and cons of both with regards to spatial and temporal dynamics. The paper is of necessity methodology heavy and descriptive, but is sufficiently clear and well-written, with potential wide ranging implications, to be of interest to the general reader of Biogeosciences.

I do have a number of concerns that require further comment and/or addressing prior to potential publication. These relate to: 1) minimum flux detection limits; 2) flux 'acceptance' criteria, and; 3) diel trends. All are important given their influence on the paper's conclusions. I discuss these initially and follow up with specific line-by-line comments. minimum flux detection limits (MDL): The MDLs for N2O and CH4 are exceptionally low – at least two orders of magnitude lower than I have seen reported elsewhere from similar automated (6-8 daily) GHG sampling

systems. Based on this alone, I have to question the approach that generates these values, and by extension the legitimacy of a major conclusion that stems from these values, ie that exceptionally low negative N2O fluxes from the wetland represent a real phenomenon. Questions arising: Is the approach used by Parkin et al. (2012) for 3-4 point manual sampling systems applicable to your presumably many hundred point automated approach? Please provide further justification.

The approach by Parkin et al. assuming 4 datapoints provides a conservative estimate of the MDL for our system with dozens of datapoints per flux measurement. We have now also included a second method which is more conservative, but which also yields low MDL estimates. The reviewer refers to other "automated" systems, and it is true that automated systems can reduce the MDL by providing many headspace concentration measurements over very short time intervals, thus increasing the degrees of freedom of the regression for the change in headspace concentration over time. However, our much lower MDL values are not only the result of automation, but mostly the result of linking the automated system to a highly sensitive laser instrument, so we have both high precision for each concentration measurement and many concentration measurements per flux estimate. Please see the response to reviewer 1 for MDL for additional explanation regarding this topic.

Does the MDL represent the Method Detection Limit or a more stringent Reliable Detection Limit (reduces expected frequency of false negatives as well as false positives)?
Were the MDLs calculated from the mean and CV of all ambient air hourly measurements over the entire measurement period at the forest site, or was this conducted periodically? If the former, you would expect a very low standard deviation and CV.

MDLs were calculated periodically and not from the overall standard deviation over the entire season. The MDL refers to a minimum detectable flux, which includes consideration of both the precision of the instrument and "the observed exchange rate and its SE for each chamber" as per Hutchinson and Livingston (1993; Use of chamber systems to measure trace gas fluxes. Pp 63-78, In L.A. Harper et al. (eds.) Agricultural ecosystem effects on trace gases and global climate change. ASA Spec. Publ. No. 55, Agronomy Society of America, Madison, Wisc.).

Would an approach that matches the air sampling time of the automated systems with that of the manual sampling system at the alfalfa site be more appropriate – this along with or separately to a randomized (Monte Carlo) sampling technique of air (or standard [see below]) concentrations would I think be more appropriate? Indeed, can ambient air be used to determine flux detection limits? Presumably N2O and CH4 concentrations vary over daily and longer time periods at both sites. Using the N2O and CH4 standards that are used to determine instrument precision and accuracy would I believe be more appropriate. Could the authors please recalculate using these standards or provide strong justification why this is not more appropriate.
In general please provide greater explanation (including equations and refs) for MDL determination, as these values are important and fundamentally impact the conclusions.

Please see response to reviewer 1 regarding recalculation of MDL using two different methods, one which includes the standard deviation of calibration gas.

2) flux 'acceptance' criteria
Using CO2 concentration change during chamber closure is a commonly accepted proxy to help to determine the legitimacy of N2O and CH4 concentrations/fluxes from the same air sample. Here, however, I question the 'toughness' of this test and the sole use of a relatively low R2 threshold for CO2 for determining whether the system was operating correctly or not. As the

authors are aware/note, although R2 is commonly reported as a means for identifying 'good' or bad' fluxes for N2O and CH4, small positive/ negative fluxes of these gases cannot be readily distinguished from zero by its use. Using R2 for CO2 therefore makes sense in this respect, but 0.90 seems a low value. The vast majority of R2s for CO2 fluxes typically encountered are >0.98 – using either manual or automated systems. Perhaps the much larger number of samples (240 over 4 mins?) per chamber closure 'depresses' the R2 value, compared with for example selecting a small number of samples, e.g. 4 (one each at 60, 120, 180 and 240 sec). Irrespective, could the authors justify the specific use of 0.9 as the R2 threshold, given that they have confidence of the previous 'linearity' of CO2 concentration changes. If the R2 values were raised to e.g., 0.95. 0.99, could the authors comment on the resulting percentage of acceptable N2O and CH4 fluxes at both sites and how (if) they alter conclusions. Although I understand the rationale behind no minimum linear criteria for CO2 with the manual chambers, other factors can of course affect CO2 concentration (e.g. improper vial evacuation, lid/septa integrity, vial storage etc). Could the authors comment on the percentage of manual samples that met the automated sampling CO2 threshold R2 of 0.9 and higher. Would a consistent approach for both systems be more appropriate –please comment?

As the number of points in a sample go up the R^2 values that are statistically significant do go down. We have previously used an R^2 of 0.90 for our automated flux sampling criteria (Savage, K., E. Davidson, and A. D. Richardson. 2008) and feel it's an appropriate acceptance level. An R^2 of 0.90 is highly significant when the number of data points in the regression (i.e., the degrees of freedom) are in the tens to hundreds, as is the case with our automated measurements. If we were to change our R^2 acceptance criteria from 0.90 to 0.95 then the % of fluxes that meet that criteria would change from 96% to 91% in Howland forest and from 90% to 81% in North Dakota.

For the manual samples there was 1 data point collected from the season that did not meet the R^2 >0.90 criteria.

3) diel trends

The correlation between diel patterns of N2O and soil temperature appears to corroborate previous studies. Blackmer et al. 1982 however note that "neither the amplitudes nor the times of minima and maxima in these patterns can be predicted solely from soil temperature". The diel patterns, particularly for N2O and CO2 may also be partially/primarily temperature artifacts associated with the warming/cooling of the sampling tube and chamber air. This appears to be corroborated by the better synchrony between air temperature patterns and N2O and CO2 diel patterns, as compared with soil temperature patterns, which have peak/trough lags of _ 4 hours (depth of N2O production questions arise?):

Although we did not measure any changes in temperature along the tubing lines, it should be noted that prior to sampling, tubing lines were flushed for 2 minutes with ambient air prior to the chamber top closing which should flush any air out of the lines that may have been influenced by changing temperatures in the tubing lines. Given the volume of the tubing lines lines (0.10 L) and the air flow rate (1 L/min) it takes approximately 6 seconds for chamber air to travel through the tubing lines to the analyzer. Further, if the tubing lines were increasing air temperatures as gas passed from chamber to analyzer, particularly in the afternoon, this would have the effect of marginally decreasing measured fluxes during this time period (for a 10^oC change this would be an approximate 3% decrease in flux).

I would therefore urge greater caution about this relationship, and reflect this in the narrative. In relation to this, what (if any) was the temperature differential between chamber air and external air during chamber closure (I realize chamber closures were very short) – was this (and any pressure differences) tested? ***We did not test pressure differences for this specific experiment however we have done extensive tests on pressure inside the chamber for this same type of system and found that with the pressure vent on the chamber top, there is no difference between internal and external pressure (Davidson et al. 2002, Ag. For.Met. 113 ,21-37).***

Specific comments
18281, L 25-26: How deep were automated chambers deployed into the peat/moss?
Did the positioning of the chambers reflect the variability of the well/poor drained site?
Were the same five automated chambers in the same conformation deployed here as in the alfalfa field (not clear from narrative)?
The five chambers were not deployed in the same semi-circular formation in the Howland wetland as in the North Dakota alfalfa field. Given the greater heterogeneity in topography at the Howland forest wetland site, the collars were placed in locations representative of the poorly drained site and not in the same pattern as at the North Dakota site. The collars were inserted 5 cm. We tested sealing between the chamber top and collar, which was manually checked on a regular basis.

18282, L 26-27: Could the authors please provide more specific dimensions for the manual sampling chambers. I cannot find chamber height, diameter, deployment depth information in the manuscript or in Phillips et al. 2009 cited in 2.3.
Collars were made of 25cm diameter polyvinyl chloride (PVC) tubing cut to 10cm lengths and inserted into the ground to a depth of 5 cm. A vented PVC chamber top was fit securely over the collars to create a closed static chamber during sampling.

18282, L 13: Were the collars inserted into soil to their full depth (5.1cm)? Given the precision needed to achieve a gas tight fit between the collars and the chamber tops using only gas pressure, what tests have been carried out to ensure no (or minimized) leakage from these systems?
We regularly check the seal between chamber tops and collars by hand and we test for more sensitive leaks by breathing at the interface between the chamber top and the collar and monitor changes in CO_2 concentration. Pressure is applied via the piston to the chamber top to help make a tight seal between chamber top and collar (Savage and Davidson 2003, J of Experimental Botany, 54, 891-899).

Can the authors describe the 'connections' used (material, dimensions, vendor etc) that allows air flow from and to the chambers and the analyzing equipment – I don't see this mentioned in the text or in Figure 1.
Synflex and bev-a-line tubing were used for gas flow and stainless steel swagelock were used for all connections.

18282, L 21: Were individual automated chambers sampled at the same time during each hour throughout the measurement periods, or where they randomized within each hour?

Automated chambers were sampled in the same order and the same time of day throughout the sampling period: sampling order or time was not randomized. We know that closing a chamber top over the soil has an effect on the gas concentration gradient within the soil and the longer the chamber is down the greater this disturbance. Randomizing our fluxes over an hourly period may result in having the same chamber sampled last in one hour, and first in the second hour with little time in between to allow soil concentration gradients to readjust, thus introducing a new source of error.

18282, L 27: Were non-linear models tested – assume short closure period precluded their use?
We did not test any non-linear models for flux analysis. In previous work on chamber artifacts (Davidson et al. 2002, Ag. For.Met. 113 ,21-37) we examined the effects of closing the
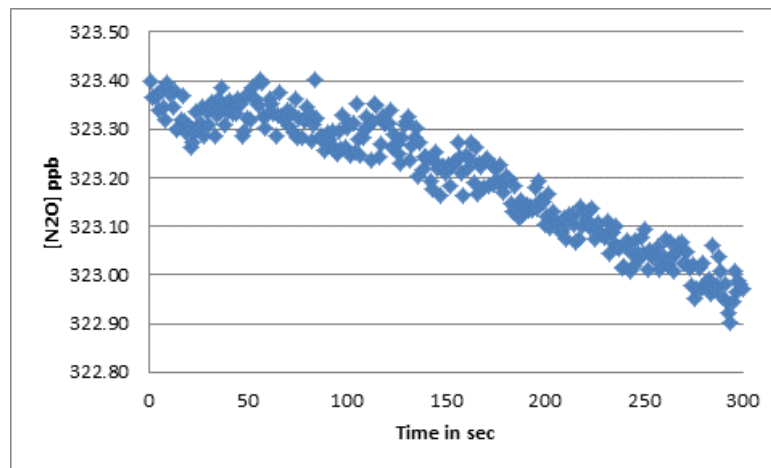

Figure 3: An example of the change in headspace [N_2O] over time from closure to end of flux calculation period. Flux calculation period is from time 60 sec to 300 sec.

automated chamber top on a collar. We found that the initial closing causes a small pulse of CO_2 to be released from the soil due to changes in pressure inside the chamber. This change in pressure is alleviated by a pressure vent in the chamber top. A non-linear curve fit to determine a flux assumes an initial undisturbed ambient concentration which is not the case in this system. We have also examined the effects of leaving the chamber top on the soil for long periods and this influence on CO_2 flux estimates and have found that our short closure period has had minimal effect on changing flux rates from soils. Figure 3 is an example of a N_2O flux trace from Howland forest wetland site, there is some indication that there is an initial short lived very small decrease in [N_2O] following chamber closure, which would be expected if lower [N_2O] gas were released from the N_2O consuming soils, however further evaluation of this is beyond the scope of this paper at this time.

18284, L 28: Was atmospheric pressure measured locally or derived from altitude?
Could the authors include a flux calculation equation for clarity?
Atmospheric pressure was measured locally at the site level. Fluxes were calculated using the following formula:

$$Flux = \frac{d[GAS]}{dt} * \frac{PV}{ART}$$

GAS represents either: CO_2, CH_4 or N_2O. Where $d[GAS]/dt$ is the change in headspace concentration of gas over time (ppm/sec or ppb/sec depending on gas), P is atmospheric pressure (atm), V is chamber volume (L), A is collar area (m^2), R is the universal gas constant and T is temperature (^oK). The flux is converted to units of either µg ($N-N_2O$ or $C-CH_4$) or mg ($C-CO_2$) and scaled to an hourly time step. This equation and example can be added to supplemental online material.

Example flux calculation for flux near MDL: slope = -1.353E^{-4}ppb/sec (0.03ppb in [N_2O] change over 240 sec), (PV/ART) = 7.50225

-0.10 µg N m^{-2} hr^{-1} = $\frac{-1.353E^{-4} \text{ mol N2O}}{1*E^9 \text{ mol air per sec}} * 7.50225 * \frac{28000 \text{ mg N}}{mol\,[N2O]} * \frac{3600\,s}{hr} * \frac{1000\,ug}{mg}$

18285, L 6: In relation to vial preparation, could the authors also comment on their practice of evacuation, as opposed to over pressurization that allows leak detection and avoids contamination during sub-sampling (eg Rochette and Eriksen-Hamel 2008.
Soil Sci. Soc. Am. J. 72:331-342)?
Vials used for manual gas sample collection (12-ml exetainers, Labco Unlimited, UK) were completely evacuated with a heavy duty vacuum pump (Welch Model 1405, Skokie, IL) no more than 1 h prior to field sampling. A pressurized sample was injected into this fixed volume to avoid contamination during gas chromatography (Rochette and Eriksen-Hamel, 2008). Samples were not stored but instead analyzed immediately following field data collection. Quality of sample vial preparation, handling and analysis was checked using standards of known gas concentrations at each field sampling.

18286, L 11: How was this particular model chosen to represent diel fluxes – were others tested? The model peaks/troughs for N2O (Fig 7a) seem shifted wrt measured N2O fluxes (earlier in morning and later in afternoon).
We have used this diel sine wave model previously for diel trends in (Savage et al 2009, Savage et al 2013) and we did not test any other models, as the parameters for this model provide us with meaningful values to test (mean flux, amplitude, and time offset) against other sites, treatments and seasonal changes. The diel GHG patterns shown in Figure 7 of the manuscript are hourly averages over the sample season, and on any given day peaks and troughs may change slightly, particularly during precipitation events. This result is meant to give an overall impression of the diel pattern for each GHG at this location and time frame.

18287, L 18-19: Please see general comments on MDL for automated system.

18289, L 13-16: Please see general comments on MDL for automated system. The range of N2O fluxes noted equivalent to _ 0.1 to 0.3 g N2O-N ha-1 day-1 are, as previously noted,

exceptionally low to consider as 'real'. Despite this, and assuming similar uptake throughout the year, could the authors comment on this sink strength at a broader scale?

The fluxes collected at Howland forest wetland in 2011 were from a short time period at the end of a sample season (fall) but given the high precision of the instrument and the low calculated MDL we are confident that the uptake measured is real.   We think that it is not appropriate to attempt to scale up or comment on sink strengths at broader scales from this limited, short term dataset for Howland Forest as it may not be representative of GHG fluxes in spring and summer. More sampling is currently underway to address that objective, but it is beyond the scope of this manuscript.

18290, L 9-12: It is difficult to see this apparent good agreement between one manual sampling event and the equivalent automated sampling event from Fig 3a-c. I suggest formalizing the correlation or consider removing/revising this assertion.

This comparison was collected at one time to show the similar range between GHG fluxes measured by the manual and automated system on this one particular date. The text is changed to reflect that this agreement only applies to one sample date.

18292, L 12-15: If manual sampling was able to "characterize much of the transient GHG responses" then how can it also "miss many of these responses". Seems contradictory, please clarify/revise.

The manual sampling of 3 times per week at the North Dakota site captured higher rates of GHG fluxes following precipitation events. The automated GHG flux system showed that the response to precipitation events can last up to a few days. Since the manual chambers were sampled 3 times per week, this frequency was sufficient to capture when GHG fluxes are high in response to precipitation events. Many manual sampling strategies are only once per week or monthly, in which case at this site they would miss many of the responses to precipitation and it is this less frequent sampling strategy we are referring to in the text when we say "miss many of these responses". Although the manual sampling at a frequency of 3 times per week captures most of the wet-up effects, the automated sampling provided more information on quantifying the pulse and its decay, and therefore provides a better quantification of this effect. We have made this distinction clearer in the manuscript text.

Section 3.6 and 3.7: Could the authors please include a few more citations in these sections, by comparing their results to a relatively small, but important number of previous publications that have investigated comparisons between automated and manual sampling systems (eg Smith et al. 2001. Global Change Biology, 7, 933-945.), determined suitable manual sampling regimes based on automated systems (eg Parkin et al. 2008 J. Environ. Qual. 37:1390-1395) and investigated diel N2O cycles and their drivers (eg Blackmer et al. 1982. Soil Sci. Soc. Am. J. 46:937-942, Alves et al. 2012. Soil Biology & Biochemistry 46 (2012) 129-135, and Smith et al 1998. Atmospheric Environment Vol. 32, No. 19, pp. 3301-3309).

Yes we can include results from these citations.

Figures
Fig. 1. As reproduced it is very difficult to read and interpret. Consider font size increases and revisions of schematic for clarity. Please also make clearer the sampling connections between the chambers and the analytical equipment in the 'enclosure'

I have increased font size and line thickness to make it more legible. I have added text to the figure caption to help clarify flows.

Fig 3. Consider removal of red symbols for manual sampling. See specific comment above.
I think that the red symbols for the manual sampling show that both types of techniques show a similar range of GHG fluxes.