Reviewer 1

Author response to BGD manuscript:

"Temporal and spatial variations of CO_2, CH_4 and N_2O fluxes at three differently managed grasslands" (bg-2013-46)

We thank the Reviewer 1 of this manuscript for their thoughtful and critical ideas towards improving this manuscript. Original comments are in regular font, and our responses in bold italics. Overall, we streamlined the original manuscript towards a better readability, leading towards the conclusions

This paper describes the compared temporal and spatial variations of soil greenhouse (GHG) gas (CO_2, N_2O and CH_4) fluxes at three grassland sites across an altitudinal gradient in the Swiss Alps from 400 m to 2000 m amsl. A general objective was to assess how soil GHG fluxes under permanent grassland respond to management and to environmental drivers (especially temperature T and soil water content SWC), with a gradient in grassland management intensity reflecting the altitudinal gradient. More specific objectives included the assessments of temporal (diurnal, seasonal, annual) variations in GHG fluxes, and of the potential errors in scaled-up fluxes associated with spatial variations among flux replicates at the field scale.

The manuscript is straightforward, well written, with clearly stated objectives and methods and an adequate description of a substantial body of flux measurements and results from the three sites. The intercomparison exercise is valuable as it could also be regarded as a global change experiment, with altered climate, temperature and soil hydrology leading to changes in both management and fluxes. However the focus here is rather methodological, and I am not altogether convinced by some aspects of the discussion of the spatial representativeness of the measured fluxes. The application of complex geostatistical methods to small datasets (N chambers = 16) rather feels like an overkill, which places the emphasis on dry statistics rather than mechanisms and process study. It may have been much more informative, for example, to measure soil moisture at each chamber location (e.g. using a portable TDR probe - if this was done, it does not say so in the text ?-), as well as available N from soil samples, which are known (or at least strongly expected) to be the primary drivers of N_2O emissions.

Response:
We agree with the reviewer 1 about the appropriateness of the geostatistical modeling and have omitted this aspect in the revised manuscript.

In addition, measuring available N from soil samples would have been one option to include soil N availability for better understanding temporal and spatial variations of N_2O emissions. Another option is to measure N concentrations of the standing biomass (what we did every month), representing a plant-integrated soil N signal, and thus leading to higher N concentrations in the plants after fertilization events. However, due to the delayed response of N transformations in the soil and accordingly into the plant, this information was omitted in the linear regression models to avoid large data losses. We added the following to the revised manuscript:"Nitrogen

inputs in the form of slurry/manure applications were not considered for the PCA, as only six and three data points would have been available at CHA and FRU, respectively, and LAI can already be seen as a proxy for management activity." (p10/l24ff)

The last figure (effect of slope on fluxes) is an attempt to quantify by proxy the impact of soil moisture, but using actual measured SWC (or % water filled pore space) values would have been much more powerful. Also, there is too litle discussion of the impact of grazing on the spatial variability of fluxes; grazing animals tend to select prefential spots for grazing (grass species composition, forage quality) and for resting (shade, wind shelter), but the manuscript does not
make it clear whether the transects were selected to yield representative sub-sections of the field (it does not appear so).

<u>Response:</u>
With our experimental setup, we covered the major management activities on each of the grasslands. This included regular harvest followed by the application of organic fertilizer, e.g. slurry and manure. Grazing as a fourth management occurred for few weeks in spring and fall at the intermediate site and for roughly 2-4 weeks at the high altitude grassland. Our setup of static chambers were not designed to account for grazing which resulted in fencing of collars to avoid their destruction. Our setup was chosen as a tradeoff between either excluding grazing by fencing the chambers or continuous disturbance of the soils inside the chambers due to removal of the chambers before grazing and their re-insertion into the soil after grazing. Therefore we chose to keep the chambers at their identical locations which excluded grazing. This resulted in an underestimation of GHG fluxes from urine and dung patches. However, this underestimation, particularly of CH4 and N2O emission was assumed less than a possible overestimation of GHG fluxes, mainly CO2, by regular disturbances caused by changing chambers at each sampling event. We clarified this aspect in the MatMet section and added a statement at the end of the revised manuscript, that recommends the use of portable chambers in conjunction with recently developed laser spectrometers allowing for much shorter sampling times and therefore sampling of additional hot spots as occurring during grazing.

Thus there is no telling whether the average of the 16 chambers is a good approximation of the field-scale ecosystem flux; the presence of hot spots (especially for N2O) makes it clear that the fluxes are log-normally distributed (in space), and thus the average could either be an over-estimate or an under-estimate.

<u>Response:</u>
We agree with the reviewer that measuring a representative flux is always tricky. However, a larger number of flux replicates naturally leads to a better representativeness of the mean flux estimate. In the literature, long-term experiments with replicated fluxes > 10 are scarce as measuring manually operated static soil chambers is very labor intensive. 16 chambers were thus a compromise between a "large-enough" number of replicates and practicability, particularly when aiming at estimating the GHG exchange for three different grasslands along an altitudinal/management gradient.

Furthermore we decided to install these 16 chambers along transects within the

(1) the footprint of the EC towers (which was less important for this study, but of large need for another research question, (2) to represent all aspects of the terrain/slope, and (3) to allow for the common farming practices. Thus, we think representing both farming practices as well as site conditions with 16 chambers at each site gives reasonable estimates of the GHG fluxes at these sites.

I also have a few reservations regarding the flux calculations and selection criteria. The paper may thus be published subject to the following revisions.

Response:

We will go into flux calculation details when replying to the specific comment below.

Specific comments

Title: Please make it clear in the text that the paper deals with SOIL fluxes only (assimilation / gross primary productivity and animal emissions are not accounted for by these (opaque) chamber measurements, and the eddy covariance data are not shown or discussed), and thus does not allow a full GHG balance.

Response:

We changed the title and all other relevant parts of the text, so it becomes clear that we were assessing soil fluxes of CO2, CH4 and N2O.

Abstract: p2636, l17: it seems rather little counter-intuitive that cattle-grazed grassland should be a net CH_4 sink, which is why it is important to state clearly here in the abstract, as well as in the title, that the fluxes shown are soil fluxes only, and thus not representative of the whole (soil+animals) ecosystem, to avoid confusion.

Response:

We agree with the reviewer and changed the title according to the suggestion made to avoid this confusion.

Abstract: p2636, l18-19: please provide annual fluxes in kg C /m2/s for CO_2 and CH_4, and kg N /m2/s for N_2O

Response:

Although we understand that annual fluxes can be of interest, we chose not to give them: In the current version of the manuscript we focus on variations of GHG fluxes at different timescales – days to weeks to months and their most important environmental drivers. Calculating annual sums from manual measurements would have called for an up-scaling/integration procedure which can be done in many different ways. In order to estimate the validity of such upscaling, additional checks would have been needed, which is actually the focus of another manuscript. An inclusion of all this info here would have led to a large increase in length of the current manuscript, making it difficult for the reader to digest the combined information.

Abstract: p2636, l24: the magnitude of the measured CH_4 uptake does not justify the use of the term "hot-spot", which is usually reserved for strong emissions. Please rephrase.

<u>Response:</u>
We rephrased this statement in the revised manuscript (p2/l20): "We found permanent hot spots for soil N_2O emissions and locations of permanently lower soil CH_4 uptake rates at the extensively managed site."

p2637, l22-25: More orographic precipitation, but also lower evapotranspiration rates due to lower temperatures, both contributing to wetter soils at high elevations.

<u>Response:</u>
We changed the sentence accordingly. (p3/l22f)

p2638, l10-12: Grazing (not mentioned here) is possibly the strongest driver of SPATIAL variability in soil GHG fluxes in grazed systems, due to urine and droppings creating emission hotspots. Grazing also drives TEMPORAL variability (l 13-15), though it could be argued that fertilisation (sudden increase in soil available nutrients, energy and N) leads to even more pronounced emission pulses than does grazing, where changes are more gradual.

<u>Response:</u>
We thank reviewer 1 for this detailed and helpful comment. Due to our experimental set-up (avoidance of disturbance by removal/reinsertion of chambers) we cannot assess the effect of grazing on GHG fluxes (see response above). However, we further included "grazing" in the respective paragraph in the revised manuscript.(p4/l9)

p2638, l21-22: "to investigate the source/sink behavior of CO_2, CH_4 and N_2O fluxes at three differently managed grasslands...": differences are not just in management, but equally important is the climatic/altitudinal gradient, whereby the effects are not easily untangled. Obviously it's the climatic gradient which has led over the decades or centuries to the development of different management practices.

<u>Response:</u>
We agree with the reviewer and are aware of the concomitant effects of both, management and altitude and therefore differing climate conditions, to GHG emissions. Therefore we added this information (p4/l21).

p2639, l10-13: was there any grazing at Chamau?

<u>Response:</u>
No, Chamau was only cut and fertilized during the period of investigation.

p2640, l10: please indicate how deep the collars were inserted into the soil

<u>Response:</u>

We added the desired information. Moreover, we included information on collar extensions, which were used for flux sampling when the vegetation was higher than 20cm, to avoid structural damage of the biomass. The new passage reads as follows:"The diameter of the polyvinyl chloride chambers was 0.3 m and the average head-space height 0.136 m (±0.015 m), average insertion depth of the collars was 0.08 m (±0.05 m). On sampling campaigns, where the vegetation inside the chambers was > 0.20 m, collar extensions (0.45 m) were used." (p6/l9ff)

p2640, l20-21: "The vegetation inside the chamber collars was manually cut at the times of regular management activities, i.e. cuts and grazing." This sentence seems to imply that the collars at the different sampling locations were fenced off - and therefore not subjected to direct grazing and cattle urine and droppings - from grazing animals? Could you please confirm? Does this then mean that the measured GHG fluxes are not representative of grazed systems, but rather of fertilised/cut grassland, and that the spatial heterogeneity normally associated with urine patches and compation by animals is not reflected in the chambers? This is an important point for the interpretation and discussion of the spatial variability later on.

<u>Response:</u>
As stated at the beginning of this document our flux measurements represent the major management activities occurring at the three grasslands excluding grazing. This exclusion is a common tradeoff when using manual static chambers. In order to avoid confusion we changed this statement to: "At FRU and AWS, chamber collars were fenced to avoid trampling and/or removal by the cattle". (p6/l20f)

p2641, l6-8: "A deployment time <40 min is considered short enough to neglect saturation effects inside the head space..." : this may have been your experience during this study, but I disagree that this can be presented as a general rule. A detectable curvature of the temporal evolution of concentrations can occur over much shorter time frames than 40 minutes, it very much depends on the source strength of the underlying soil, but also on the collar depth. Theoretically the concentration in the chamber will level off at the concentration occurring in soil pore space at the depth of the base of the collar.

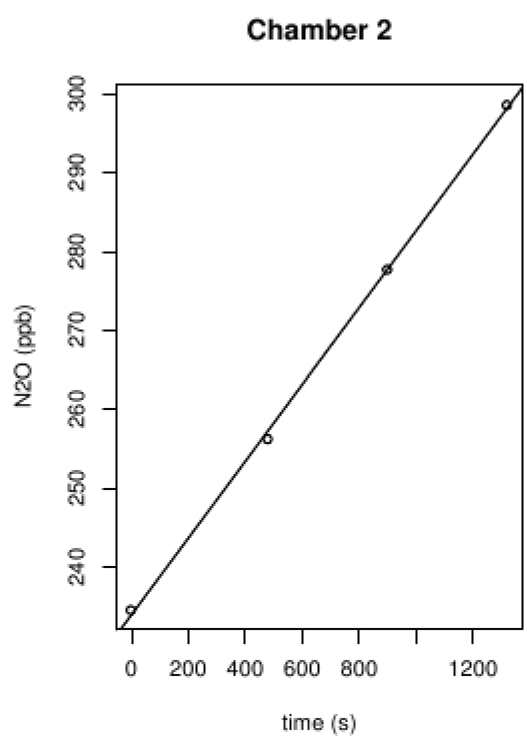
<u>Response:</u>
We agree with the reviewer that our statement can not be taken as a general rule. Deployment times should always be adjusted to the magnitude of the expected flux the head space volume as well as the chamber type (ventilated/non-ventilated) amongst other factors. In our specific case we never observed saturation effects, neither after slurry applications at the intensively managed site nor during wet periods at the research site with moderate management. We changed the sentence to: "This closing time was sufficiently short to avoid saturation effects inside the chamber head spaces." (p7/l8)

p2641, l19-20: "... calculated by the slope of the linear regression between gas concentration and time. Fluxes were always small enough that no curvature in measured concentration data could be detected which would be indicative for saturation effects inside the chamber": How did you quantify the curvature? Using the linear Rˆ2 ? I would argue that no curvature was detected presumably also because only 4 samples per chamber flux measurements were taken, such that the noise to signal ratio over the 40 minutes was comparably high. Chamber measurements made using fast response continuous gas analyzers (e.g. LiCor or other IRGA for CO2) have demonstrated that at least a slight curvature can be almost always detected, and is indeed theoretically expected, resulting in systematic flux underestimation (typically 10-30%), even though the linear Rˆ2 is consistently and comfortably well above 0.98. A very useful reference is Petersen et al. (2010), A comprehensive approach to soil-atmosphere trace-gas flux estimation with static chambers, European Journal of Soil Science, 61, 888–902. An alternative re-computation of your fluxes using the Petersen et al. algorithm in R would be very informative, providing an uncertainty estimate for individual flux measurements.

Response:

We thoroughly revisited our flux calculations. Yes, we quantified the curvature using the R2 of the linear regression between concentration changes over time. In addition, we computed fluxes using the intercept method according to Kroon et al. (2008). This methodology is also aiming to avoid underestimation of fluxes due to non-linearity over time. We however found, that resulting flux estimates of N2O and CO2 (both, were expected to show increasing concentrations with time) were not much different from the flux estimates obtained by the simple linear regression approach we used. For N2O, computed fluxes (intercept method) were on average 7.9% smaller as compared to fluxes calculated with linear regression. For CO2, respiration would have been 11.3% larger.

Furthermore we visually inspected the slope between the concentration measurements between time steps 3 and 4, especially for N2O after fertilization events at CHA (e.g. Fig. below). We found that for N2O, the slope was in ca. 50% of the cases the same or even steeper as the slope between the previous concentration measurements (time steps 2 and 3 and time steps 2 and 1). In addition, we computed fluxes using the intercept method according to Kroon et al. (2008). This methodology is also aiming to avoid underestimation of fluxes due to non-linearity over time. We however found, that resulting flux estimates were not much different from the flux estimates obtained by the simple linear regression approach we used. Therefore, we chose to rely on fluxes that were calculated by the linear regression approach only.



p2642, l1-2: "...fluxes were only computed if the linear regression yielded a r2 >0.8." Do you mean by this, that GHG fluxes (for the three gases CO2, N2O and CH4) were only computed when the Rˆ2 for CO2 was above 0.8, with CO2 being taken as a quality criterion for the whole chamber operation (based on the assumption that there must always be CO2 mineralisation and thus evolution from the soil, and that any noisy CO_2 temporal profile indicates a dysfunction of the sampling system, for example a large leak)? Or do you actually mean that for each gas taken separately you applied a selection based on the R^2 of the gas in question? In the case of the latter, consider the hypothesis that the N_2O concentration is almost exactly 320 ppb (+/- the uncertainty in the GC concentration measurement) at sampling times t0, t10min, t20min, t30min. The R^2 is very close to 0, and yet the flat concentration profile tells us that net emission or uptake takes place. Discarding all such flux events would inevitably bias the temporal or spatial average upwards (overestimation of the annual-scale and field-scale emissions). Please comment.

Response:
We applied the R2 threshold per GHG, and not only for CO_2. And in the case of CO_2, our concentration increases were in 92% of the cases linear with R^2>0.96. We would like to thank reviewer 1 for the additional comment concerning omitting flux values due to low r^2 values of the linear fit. In general we agree with the fact of counteracting processes. However for N_2O these counteracting processes are very difficult to disentangle and to our knowledge few studies have reported considerable N_2O uptake rates. Furthermore N_2O uptake and the involved processes are still unclear.

In our study N_2O concentration changes (t0-t30min) were greater than ppb in 90% of the cases. Especially after fertilization, we observed very large concentration changes over time (e.g.>120 ppb and still not saturated as the increase between t3 and t4 was larger than between t2 and t3)

We however added the following information to the revised manuscript: "If the slope between the first and second concentration obviously deviated from the one of the remaining three concentration measurements, we omitted it and calculated the flux from the remaining three." In total this happened in 32% of the N_2O fluxes. In any case omitted data was commonly based on single chambers only.

In addition we included a comparison between chamber and EC derived soil fluxes of CO_2 in the discussion paragraph of the revised manuscript (p16/l7).

p2642, l2-4: "CO_2, CH_4 and N_2O fluxes per chamber which were then filtered for obvious out of range values (±10 SD) for each sampling campaign": on what statistical population is the SD computed, on the 16 chambers measured on each day at each field? Does that mean that you discard and exclude hot spot fluxes from the spatial average, thereby artifically reducing both the natural variability in fluxes and the arithmetic mean emission or uptake? I don't understand the rationale. As long as each chamber flux has successfully passed the individual flux selection criteria (based on R^2 as discussed above), I don't see any reason for rejection, especially for gases like N_2O and CH_4, which are known to vary spatially/temporally over several orders of magnitude.

Response:
We agree with the comment given by reviewer 1 and omitted this filter during the reanalysis of the data which is now included in the revised manuscript. All valid individual chamber fluxes, where the linear regression for the concentration changes over time showed r^2 >0.8, were used. Please note, that

there is no difference in the presented flux data notable as very few data points were excluded via this filter originally.

p2642, l17-22: What do you mean by each "site"? Do you mean at each "field site" (CHA, FRU, AWS), with one soil measurement station for T and SWC for the whole field? Or did you measure T and SWC at each of the 16 "flux chamber/collar sites", not necessarily continuously but at the time of each flux sampling campaign, using portable soil probes?

<u>Response:</u>
With "site" we original meant "field site", and changed this in the revised manuscript accordingly. We measured air and soil temperature at at one point for the whole grassland. We added this information to the MatMet section: "At each field site, the following environmental variables were recorded at one point (eddy covariance tower) as 10 minute averages:" (p8/l15ff)

p2643, l4-12: only linear regressions are mentioned in the statistical models fitted. Yet the temperature responses of at least CO_2 fluxes (soil/ecosystem respiration), and possibly N_2O in wet soil conditions, are expected to be exponential, not linear (cf Lloyd and Taylor, Funct. Ecol., 8, 315–323, 1994, functions for CO_2, on which countless gap-filling exercises in the EC flux literature have been based). Similarly, for N_2O, the SWC response is expected to be bell-shaped, with optimum conditions for nitrification and denitrification occuring at mid-range and upper mid-range (around field capacity), while at saturation (100% water-filled pore space) denitrification proceeds all the way to N_2, and thus the N_2O efflux decreases; thus the SWC response can't be linear. It seems to me that the fraction of the (temporal) variance in GHG fluxes you are trying to explain using multiple regression approaches can only be underestimated by using strictly linear functions, and that a more process-based selection of models would benefit the analysis. (see e.g. Meda et al., Biogeosciences, 9, 1493–1508, 2012, in which we also studied spatial and temporal variations in GHG fluxes, in that case of the grass outdoor run of free-ranging chicken; or again Luo et al., Biogeosciences, 9, 1741–1763, 2012, Decadal variability of soil CO_2, NO, N_2O, and CH_4 fluxes at the Hoeglwald Forest, Germany).

<u>Response:</u>
We tested both, linear and exponential regression models (uni- and multivariate). In the cases of CO_2 and N_2O, linear models always explained more variance than the exponential models. The only exception was found for CH_4 at Früebüel and Chamau, where a larger variance (9 and 16%, respectively) was explained than compared to a linear fit. In order to keep the set of methods/models consistent among the sites and greenhouse gases, and to assess the relative importance (from hierarchical partitioning) of respective drivers, we used multiple linear regression models at all sites for all gases. According to Chevan et al. (1991), the hierarchical partitioning method for the estimation of the relative importance of a single driver within a multiple linear regression model is clearly dominant over simple methods (e.g. beta-values). Therefore, we decided to log-transform CH_4 data from FRU and CHA and to use this as an input for the multiple linear regression models. We added the following sentence to the M&M section: "Since CH_4 fluxes at CHA and FRU showed exponential relationships with soil water content, data were log-transformed for the respective multiple linear models." (p9/l5)

For N2O and SWC, we found hardly any systematic relationships. We are aware of the publication by Castellano et al. (2010), who found bell shaped (Gaussian) functional relationships, with peaks of N2O emissions at different WFPSs. However, there are other studies, which report linear (Itoh et al. 2012) and exponential relationships (scholes et al. 1997), or no relationships at all (Agnew et al. 2010). This indicates the variety of responses of N2O fluxes to SWC, which might be caused by the fact, that at some soils WFPS > 80% (the range at which denitrification becomes dominant) is hardly reached.

p2644, l7-8: "CH4 fluxes did not follow any seasonal trend": there does seem to be systematic seasonality in CH4 fluxes, with Fig. 2 showing consistent uptake in spring-summer-autumn (Jul-Nov and then Mar-Jul), and consistent emission in winter (Nov-Feb) at both CHA and FRU.

<u>Response:</u>
This sentence was changed to: "CH4 emissions were mostly observed during winter, whereas uptake rates were prevailing in summer."

Incidentally, the figure caption in Fig.2 should say "...mean SOIL flux of the respective greenhouse gas...", not "ecosystem" flux, for reasons explained above. The rest of the document should be checked for similar occurrences.

<u>Response:</u>
Done.

p2644, l19: "At AWS, an average efflux of 0.23 nmol m−2 s−1 was observed": were there any measurements during or just at the end of the melting of the snowpack, and during freeze-thaw events in late winter and early spring, which could both release substantial quantities of N2O?

<u>Response:</u>
Since the site is inaccessible during winter due to avalanche danger and snow cover a large lack of data is present. First measurements in 2011 were well after possible freeze-thaw emissions. A recent study by Merbold et al. (2013) focusing on GHG emission, including N2O, from a sub-alpine grassland in winter did not report such peak emissions.

p2644, l25: what kind of fertilizer response of CH4 fluxes would you have expected?

<u>Response:</u>
Our hypothesis were lower uptake rates of CH4 due to the decreased oxidative capacity of the methanothrophs after fertilization,

p2645, l13: please change "small-scale GHG flux variability" to "temporal GHG flux variability", to make it clear that this paragraph is not about the spatial variations. My comment here again would be that the PCA based on linear models might miss important non-linear features such as the effect of SWC on N2O (e.g.)

Response:
Done

p2646, l12-13"SWC had much less influence on the N2O efflux, with a RI of 16.1 %"; the linear approach may completely miss the increase of N2O emissions at field capacity due to the reduced emissions at saturation?). Linear models explain 19-42% of temporal N2O variations, and it would be interesting to see whether non-linear models can show higher explanatory power. Also, I wonder why management was not included as additional factors in the analysis to explain the temporal variations, as surely the introduction of grazing animals and fertilisation should be strong drivers of seasonal changes in fluxes.

Response:
As previously stated, our date were checked for SWC influence on N2O fluxes, without leading to a clear result and we found no relationships. This result was also corroborated by a follow-up short-term experiment in which we assessed short-term variations in WFPS on N2O fluxes. Our findings are in agreement with a study by Agnew et al. (2010) whom neither found N2O fluxes systematically responding to changes in WFPS.

p2648, l7-25: please specify what metric is used to quantify "spatial heterogeneity" to compare between sites and gases. Presumably the coefficient of variation defined by CV=SD/average?

Response:
We used the range of the 95% confidence intervals of the mean fluxes of the respective GHGs as a metric to evaluate the spatial heterogeneity. This nicely shows the spread of individual chamber fluxes and thus elucidates the spatial heterogeneity of GHG fluxes.

p2648, l8-9: "Spatial heterogeneity in N2O fluxes was largest at AWS...probably due to large variations in SWC": but also possibly due to a larger impact of grazing, since the other 2 sites were also cut and thus the grazing fraction of time was reduced compared with AWS?

Response:
We were unable to assess grazing effects within this study, as chamber collars were fenced and thus not subject to feces or urine.

p2649, l9, or p2643, Statistics section: please define "spatial auto-correlation" mathematically, as well as "semivariograms", for the lay reader.

Response:
We no longer present results regarding the semivariariograms or spatial auto-correlation.

p2653, l18-19: you argue that significant diurnal changes in N2O fluxes were observed at FRU because the site had recently been fertilised (contrary to CHA). I do not find the argument entirely convincing; if the diurnal process is temperature driven, there is no reason why the relative effect, as quantified by the % change (day-night)/day, should be different. Unless what happens is that, in background situations (long after any fertilisation has taken place), the measured N2O fluxes are not significantly different from zero for both day and night (i.e. within the uncertainty of the flux measurement system, because of random errors in concentrations). If fluxes could be measured (with a more precise flux system) then one might also see diurnal flux differences.

<u>Response:</u>
We agree with the statement of reviewer 1 on the importance of more precise flux measurements after management events. We think, that simply not enough nitrogen was available at CHA to form significant diurnal changes in efflux, as the last fertilization was almost four weeks prior to the intensive sampling campaign. Thus, we think that N2O emissions reached the background level of efflux, which simply is so small that temperature exerts a non-measurable influence on N2O flux magnitudes. This is corroborated by recent high-frequency observations at CHA, which do not necessarily show diurnal variations three weeks after fertilization.

p2654, l2-4: "Working with soil chambers requires information on the spatial distribution of GHG fluxes at ecosystem scale to design appropriate experiments and to be able to correct mean ecosystem fluxes for potential biases". Does this imply that the flux variability, and thus the fluxes themselves, should be known a priori (before the measurements start)? Obviously this is not possible, but one recommendation may be that the field be divided into several sub-sections assumed to be homogeneous, based on objective criteria: terrain slope (as rightly noted p2655, l5-10); soil moisture mapping; preferential grazing and resting areas for animals; vegetation species composition; soil hydromorphy; etc.. Each section should ideally be sampled with a number of chambers proportional to the expected emission rate, because hotspots dominate the field scale emission and thus the measurement effort should be commensurate. However, in practice, few studies can afford such intensive measurement efforts.

<u>Response:</u>
We fully agree that such studies assessing the site characteristics prior to flux sampling are not feasible. We however think that on sloping terrain the locations of the chambers should adequately represent all aspects of the terrain and the species composition. This does not require labor intensive pre-studies, but only a thorough visual inspection and consideration. We added the following to the revised manuscript:"Thus, on sloping terrain, mean chamber fluxes of CH4 should be estimated from an ensemble that is (a) big enough, (b) adequately represents species composition and thus grazing hot-spots, as well as (c) the average slope of the site." (p18/l15ff)

p2654, l12-13: "Chambers placed in terrain with greater inclination systematically exhibited lower SWC values." Was this a visual observation?

<u>Response:</u>
This statement was based on our automated regular measurements of SWC at several locations. Since these measurements did not cover the full period of flux measurements, we did not include this data in our analysis. We changed this passage to:"Chambers placed in terrain with greater inclination systematically exhibited lower SWC values (data not shown)." (p17/l21f)

p2654, l17-18: "Omitting permanent hot spots may lead to a systematic bias in GHG flux budgets." This is true; however, an over-representation of hotspots will conversely lead to an over-estimation of GHG emissions; and so the question of what fraction of chambers should be hotspots remains entirely open. Thus the spatially averaged GHG fluxes presented in this paper, as in most others, may be over- or under-estimates of the true field-scale integral.

<u>Response:</u>
We agree that we are not able to state whether we capture true emission or uptake rates that occur at the AWS site. Our intention was that neglecting certain aspects of exposition in sloping terrain will further lead to biased flux estimates, especially when soil water conditions are the dominant driver, as in the case of CH4. Therefore we added: "Thus, all aspects of exposition and slope should be covered when assessing flux estimates of CH4 and N2O in sloping terrain."

p2655, l5-10: the recommendations for sloping terrain are useful, but should be more general and include other considerations about local-scale drivers of spatial variability, such as mentioned above. And what is "big enough"? I would argue that a workable compromise should be based on 1- routine measurements using a few (N = 5 - 10) automatic or manual chambers with adequate temporal cover, and 2- a few campaigns in each season with high spatial resolution measurements using a "fast-box" system (Hensen et al., Agriculture, Ecosystems and Environment 112, 146–152, 2006) to map out the spatial variability and thus extrapolate fluxes to the field scale.

<u>Response:</u>
We chaged the revised manuscript to: "Thus, on sloping terrain, mean chamber fluxes of CH4 should be estimated from an ensemble that is (a) sufficient in size, (b) represent the common species composition including hot-spots occurring due to grazing, and (c) the terrain of the site. This is important since SWC is one of the major environmental drivers of CH4 exchange." (p18/l15)

Technical corrections
p2641, l24: concentrations are actually mixing ratios and their units are µmol/mol (ppm) or nmol/mol (ppb)

<u>Response:</u>
We corrected this.

p2651, l27: "envIronment"

<u>Response:</u>
We corrected this.

p2652, l9: "..taken IN generally drier soils..."
<u>Response:</u>
We corrected this.

p2654, l9: change "merely" to "only"

Response:
We corrected this.

p2654, l11, and Fig.9: suggest change "inclination" to "slope"

Response:
The term slope may transport information on the exposition, thus we kept using inclination at this point.