

Reply to Anonymous Referee #2

Dear Madam, dear Sir,

Thank you for your careful reading – your detailed suggestions helped us to improve the MS considerably.

We acknowledge your justified statistical considerations about the replication of fertiliser treatments, but we do not agree with your suggestion to declare all testing between the respective treatments as invalid. Here is why:

The discussion about pseudo-replicates exists since there are ecological field studies. A true replicate in a statistical sense is independent of the other replicates for a factor level, treatment, or whatever. When sampling in the field, the question arises “independent in regard to what”? Shall the climate be independent (because climate variation is a strong driver of GHG production and emission)? But then replicates would have to be placed many miles apart, which is rarely the case. Shall the soil be independent, i.e. should the measurements be carried out in randomized, not neighbouring blocks of a block design? But then, how to handle studies without treatments? Here, it is often assumed that the sampling locations are independent when they are placed far enough from each other. But how far is far enough? Well, that depends on the study site itself and for our study site we have good indication, that our sampling locations are independent enough to treat them as replicates because the soil of the study site shows very high small-scale variability well below the meter scale which tends to level off with increasing scale at the site level (Jacobs, 2014; LFA, personal communication, 2012). This is a strong argument for our results won't being any better or “more” valid if we had positioned the single sampling locations in greater distance from each other, or had spread them across gas sampling locations. Especially the relatively high standard deviances of our soil data corroborate this assumption (e.g., Py-FIMS data in Table 2). Furthermore, pseudo-replicated time series are common for soil gas measurements with chambers, since the bases for chamber deployment are anchored into soil to minimise soil disturbance (cf., Clough et al., 2015; Parkin and Venterea, 2010). Also, trace gas fluxes exhibit a high degree of spatial variability and it is, therefore, recommended to use rather chambers with large footprints (as ours) to integrate this variability (ibid.). In other words, the placement of our bases and their distances to one another were appropriate.

According to Schank and Koehnle (2009), it is always necessary to consider the objectives of the study when discussing pseudo-replication. Our study aims at estimating gas fluxes (CO₂) from soil immediately after tillage to link them to short-term changes of SOM analysed by Py-FIMS. Based on this objective, we had very strong logistical reasons to carry out the sampling as described in the MS. First, the adequate coverage of these highly dynamic flux changes requires high frequency measurements in terms of minutes. This would have been hard to implement with sampling locations sharing greater distances. Second, analyses by Py-FIMS are rather expensive (approx. 600€/sample). Therefore, it was beyond our logistical and financial opportunities to gather data from all four replicates (plots) of each treatment. Thus, we decided to concentrate our resources and efforts on one plot of each treatment for these analyses.

After all, pseudo-replication is only a cause for concern when the results are used to generalise outside the study system, which means that generalisability of pseudo-replicated studies may be

lower than for those with true replication (Haddaway et al. 2014). We acknowledge the pseudo-replication issue now in the methods and conclusion sections and argue as outlined above:

- 'Since the soil of the study site shows high small-scale variability (LFA, 2012, personal communication), sampling locations were expected to be independent enough to be treated as real replicates. Especially the relatively high standard deviances of our soil data corroborate this assumption.' (now 204-207)
- '...SOM composition of the investigated soil changed in the temporal scale of days...' (493-494)

We agree that the discussion about the link of carbon use efficiency to microbial activity may be too speculative and removed the related text, accordingly.

In the following, we answer to your specific comments. In parentheses at the end of our answers we give the lines in which the text is now.

37: Rather: '...by the disruption of macroaggregates, leading to release of protected SOM. . '

- Changed accordingly (36)

44: Remove author initials in references

- We removed the author initials.

52: The way the link between SOM and SOC is expressed sounds odd also in the reply to R1 ('Admittedly, SOC accounts for the majority of SOM, but. . .'). And maybe there is no need to introduce this issue here (but see Pribyl, 2010). To simplify, I would rather suggest to write: 'However, these correlations do not causally explain which organic components are mineralized.'

- Suggestion accepted (50-51)

88: Although with some updates, this section (88-100) is closely following Fiedler et al. (2015), and some condensation would be appropriate, e.g., replacing 91-95 by: "The top soil (0-30 cm) had an organic C content of $1.16 \pm 0.1\%$, pH of 7.4 ± 0.9 , and bulk density of 1.51 ± 0.08 (mean \pm standard deviation, $n = 3$) as measured according to Fiedler et al. (2015)." Also, for such an important parameter as pH for mineralisation, it would be reassuring (and easy) to provide more robust information than pH 7.4 ± 0.9 . As a detail, always give same no. of decimals in mean and measure of variability (i.e., adjust $1.16 \pm 0.1\%$).

- The text in lines 91-95 was replaced according to your suggestion: 'The top soil (0-30 cm) has an organic carbon content of $1.16\% \pm 0.10$ (mean \pm standard deviation, $n = 3$), pH of 7.4 ± 0.9 ($n = 3$) and bulk density of $1.51 \text{ g cm}^{-3} \pm 0.08$ ($n = 3$), measured according to Fiedler et al. (2015).' (97-99)
- You are right, the pH should be measured preferably from every soil sample in studies like the present one. We will consider this in the future.

108: As the BD was applied to 160 kg N/ha in 2012, it would seem the assumption of 70% available N in total N was founded before the personal communication in 2014. The assumption is ok, but the writing could be made more consistent.

- Yes, the assumption was obviously founded before fertiliser application, corrected to 2012 (112).

112: What is meant by 'in original matter'?

- Thank you for mentioning this. We checked back again and found, that this was a wrong translation obviously. Instead, we meant 'undried material' and changed the text accordingly (116).

112: Suggest completing the chronological description: 'During the cropping season 2012, maize was grown according to conventional agricultural practice.' (and add if any special management were implemented)

- We changed the text according to your suggestion (118-119).

118: Give dimension (b x w) for collars

- 'The bases had dimensions of 79 x 79 cm' (127-128).
- Additionally, we changed 'collars' to 'bases' throughout the manuscript.

120: Maybe better: 'The adjacent collars were placed 1 m apart.'

- We changed the text according to your suggestion but used "bases" instead "collars" (see above) (116-117).

128: Yes, even though acknowledged, pseudo-replication is of course a weakness of the study. And the HSD statistics as shown, e.g., in Table 1 are severely compromised by this design and should be reconsidered and preferably omitted

- As pointed out above, the placement of the bases and, hence, the pseudo-replicated time-series are common for trace gas measurements and in our view are appropriate in regard to spatial variability. That's why we think that the statistics are not compromised.

141: Suggest splitting sentence to improve readability: '... of a concentration measurement. This allowed discarding data...'

- Changed accordingly (146)

159: Does this literally mean 5-15 cm depth or is it 0-5 to 0-15 cm? Please specify. In the former case, why was the top-soil not considered?

- Thank you for bringing this up. We thought again about it and think that this information was an unjustified simplification. Here is why and our solution: Due to the uneven soil surface after tillage, it was not possible to sample accurately at 0 cm. Therefore, we decided to write "5 cm discarded" in the first place. After careful consideration, however, this does not reflect the real depth discarded. Therefore, we corrected the phrase and specified also the height of the used sample rings: '...were taken between 0 – 10 cm depth (depending on unevenness of soil surface due to tillage) with soil sample rings (h = 6.1 cm, V = 250 cm³) ...' (164-165).

218: For N only partly true and for HWN claim not statistically supported by data (even with faulty HSD)

- Maybe this phrase was a bit ambiguous with regard to the comparison between treatments. Now, we point out the relation between MF and BD: 'Before tillage, the soil of all treatments had similar C and HWC contents, but differences appeared between MF and BD, where the N and HWN contents were slightly, though not significantly, higher in MF, resulting in significantly narrower C/N and HWC/HWN ratios in MF...' (242-245).

166: Since 5 mg samples were used, a full description of procedures for homogenisation and subsampling should be given

- We added more details according to your suggestion and the text now reads: '...the freeze-dried samples were finely ground and homogenized by a planetary ball mill. Then, about 2 g were transferred into a Petri dish with a spatula and three crucibles were filled by drawing them across. These subsamples of about 5 mg were thermally degraded...' (171-174).

191: What is meant by 'total nitrogen bound' ?

- This term refers to the sum of the organic and inorganic bound nitrogen, with the exception of elementary nitrogen. We adjusted the respective phrase: '...determination of hot-water extractable organic C (HWC) as well as of organic and inorganic bound N, often referred to as "total nitrogen bound" (HWN). These measurements of organic C and total nitrogen bound are based...' (198-199).

273: '...but it was not significant ($p < 0.01$).' Correct p value ?

- Yes, indeed, another typo. Please, apologize. The correct value is $p > 0.1$ (298).

298: rather '...in contrast to...'

- Changed according to your suggestion (329)

299: '...to phenols, lignin...' or rather '...to phenols and lignin...' i.e., as individual compounds or as a group?

- No, not as individual compounds but as a group. We tried to make this more clear by editing the text to: '...to phenols and lignin monomers as well as to heterocyclic N-containing compounds...' (329).

301: Correlation to free fatty acids not shown in Fig. 6 (free fatty acids not included in figure)

- Thank you for your exact observation. We added the respective row to Fig. 6.

306: Don't abbreviate Table

- Changed accordingly

306: See comment to 218

- We changed the text and added information according to your suggestion: 'The C-, HWC-, N- and HWN-contents of the treatments showed no significant differences before tillage (Table 1), thus confirming the outcomes of other field experiments with similar fertilisers (Makádi et al., 2016; Odlare et al., 2014). However, the C- and N-contents obtained may not be

representative for long-term effects of biogas digestate vs. mineral fertiliser which may also depend on soil texture (Makádi et al., 2016).’ (337-342)

322: Speculative argument (322-323)

- Yes, we know, but we would like to discuss the issue anyway. However, we attenuated the argument by changing phrasing from ‘most likely’ to ‘possibly’ (351).

344: No need to invoke such ratios; suggest to delete 344-346.

- Deleted according to your suggestion.

348: Less available than what? You don’t have a reference with the native biomass. Indeed this section is too speculative and should be shortened (347-365).

- Less available than undigested matter, which is generalizable (cf., Möller, 2015), and thus, there should be no need to provide a reference. We changed the phrase to make this clearer for the reader: ‘... likely less available to soil microorganisms than undigested organic matter, i. e., more “recalcitrant”, since the most labile C is generally consumed in the biogas reactor ...’ (374-376)
- In addition, we omitted the speculation about C use efficiency induced by N availability: ‘As a consequence, the effect of increased CO₂ efflux after tillage as observed in CL and MF, may have been substantially reduced by a relative shortage of labile substrate for soil respiration in BD. The proportion of carbohydrates in BD derived from Py-FIMS, as discussed below, indicates not limited, but rather low levels of available C in the soil of BD.’ (379-384)

370: Leinweber et al. (2009) is not included in reference list

- Thank you for this comment. We now added the reference and checked back through the whole reference list if all citations are in there.

370: ‘complementarily’ – do you mean, e.g., ‘collectively’

- Yes. We changed the wording accordingly (389)

385: be specific: ‘The VM increase. . .’

- Changed accordingly (403)

442: ‘Since microbial activity can affect heterotrophic respiration. . .’ this seems to be an understatement; delete ‘can’

- We deleted “can” according to your suggestion (470)

449: Again, this add to the many speculations put forward in the discussion (449-451)

- We deleted the sentence according to your suggestion.

455: I don’t think the data has shown that lower CO₂ efflux was related to higher C use efficiency?

- Yes, as discussed above, we removed the speculation about C use efficiency. However, since the outcome of the PLSR supports our suggestion of enhanced microbial N-turnover, we

substituted the respective phrase by 'lignins, ammonia and ammonium were discriminated as explanatory variables for cumulated CO₂ efflux by PLSR' (480-481)

463: An awkward formulation: 'which likely inhibited an enhanced microbial activity.' Maybe use 'prevented' rather than 'inhibited'. Anyway the C/N ratios shown in Table 1 seem not to be critical for microbial mineralisation; rather N mobilisation would occur from mineralization.

- We replaced 'inhibited' by 'prevented': 'the absence of amendment may have led to a relative shortage of labile N as indicated by the higher HWC/HWN-ratio which possibly prevented an enhanced microbial activity.' (486-488)
- Though the C/N ratios might not be critical, the HWC/HWN could be "less optimal" in terms of availability of readily decomposable N. Recently, the availability was acknowledged as crucial for N use efficiency by Andresen et al. (2016). This is also discussed in 443-447, now: 'minor changes in SOM compounds in CL might be a consequence of the wider HWC/HWN ratio compared to MF and BD since it indicates a lower availability of labile N for microbial utilisation (Mengel, 1996). However, the total C/N ratios were not critical for microbial activity (Table 1) (Kuzyakov et al., 2000).'

Reference list is somewhat excessive (74 refs) with 44 unique references in the discussion alone, underlining the need for a more stringent focus in the discussion.

- We carefully went through all citations and omitted 27 unique references which provided no additional value (but also added a few during revision ...).

666: Ohkubu et al. (2012) not cited in text

- In the process mentioned above we also deleted references that were left-overs from previous versions of the MS, like this one

560: add editor info

- Added accordingly (578)

567: for consistency spell out journal names (also 590, 647, 565, 709, 736)

- Changed accordingly

Table 1: Define in Table caption the meaning of pre, post and post+4

- Changed accordingly

Table 2: As for Table 1

- Same here

Figure 1: Spell out rev. plough in caption

- Changed to 'Mouldboard plough'

Figure 2: ‘. . . (harrowing to ~ 10 cm depth and ploughing to ~ 30 cm depth)’ ‘. . .in order to better visualize the tillage effect’ Indicate specifically what boxes, error bars, points and symbols refer to in this case

- Boxes and symbols for fluxes on days after tillage are now ‘distinguished by light grey backgrounds’

Figure 6: This design is not helpful at all to the reader; please indicate R values numerically rather than by area-based symbols. And add explanation of acronyms.

- R values are added
- For explanations, to retain flow of reading, we referred to Table 2: ‘(for explanation of abbreviations see Table 2)’.

References

- Andresen, L. C., Björsne, A.-K., Bodé, S., Klemetsson, L., Boeckx, P., and Rütting, T.: Depolymerization and mineralization - investigating N availability by a novel ^{15}N tracing model, SOIL Discuss, 2016, 1–21, doi:10.5194/soil-2016-11, 2016.
- Jacobs, O. (2014). Vergleich von zwei unterschiedlichen Kammersystemen zur Messung von bodenbürtigen Lachgasflüssen. Master thesis, University of Rostock
- Clough, T. J., Rochette, P., Thomas, S. M., Pihlatie M., Christiansen, J. R., Thorman, R. E. (2015). Chamber design. In: de Klein C. A. M., Harvey M. J. (eds.) Nitrous Oxide Chamber Methodology Guidelines, Version 1.1 Ministry for Primary Industries, Wellington
- Haddaway, N. R., Burden, A., Evans, C. D., Healey, J. R., Jones, D. L., Dalrymple, S. E., & Pullin, A. S. (2014). Evaluating effects of land management on greenhouse gas fluxes and carbon balances in boreo-temperate lowland peatland systems. *Environmental Evidence*, 3(1), 1.
- Möller, K.: Effects of anaerobic digestion on soil carbon and nitrogen turnover, N emissions, and soil biological activity. A review: *Agronomy for Sustainable Development*, Agron. Sustain. Dev., 35, 1021–1041, doi:10.1007/s13593-015-0284-3, 2015.
- Parkin, T. B. and Venterea, R. T. (2010). USDA-ARS GRACEnet project protocols, chapter 3. Chamber-based trace gas flux measurements. *Sampling Protocols*. USDA-ARS, Fort Collins, CO, 3-1.
- Schank, J. C., & Koehnle, T. J. (2009). Pseudoreplication is a pseudoproblem. *Journal of Comparative Psychology*, 123(4), 421.