Dear Editor,

Thank you for commenting on our manuscript. Most of your comments are very well received and definitely help to improve the manuscript. However, we have the impression that you based your recommendations on an older version of the MS because many of the suggestions we already acknowledged and changed the MS accordingly. Below we provide detailed responses to all your comments including how we addressed it in the most recent version of the MS that is uploaded to the interactive discussion section (supplement to AC3).

Regarding the question of "pseudo-replicates" we provide detailed responses below. However, we want to express the following upfront: You know that most published work based on "cutting edge" methods with "objectively" very limited numbers of samples that can be run (NMR, mass spec, synchrotron) has almost no replications (neither true nor pseudo). In THIS context our data base is sound and there is no reason for us to withdraw any conclusion and we think that for a first time analysis with such an expensive analysis method, the sampling design is sufficient.

Thus, we look forward to a positive decision very soon.

Topical Editor's report:

Comments on “Tillage-induced short-term soil organic matter turnover and respiration” by S. R. Fiedler et al.

I agree with the comments of the two reviewers and in general also with the response of the authors. I would recommend major revisions of the manuscript before it might be considered for publication.

However, I have some further comments / suggestions for the revision of the manuscript:

I have similar problems as one of the reviewers with the statistical design of the study. The authors should be very clear that they used pseudo replicates. If I understood the paper correctly, the authors used just one experimental plot (6 x 30 m) and installed the three collars for CO2 measurements just in a distance of 1 m. This very small area was used to take the three soil samples too. This sampling strategy is not really representative for a field experiment also taking the spatial variability into account. The advantage of this strategy could be that small temporal changes might be detectable because of reducing the spatial variability.

- Our study’s aim was to get a first insight into the short-term SOM turnover after tillage and, thus, into the temporal changes which, indeed, were obviously detectable. From this results we confirm that Py-FIMS is an appropriate method to study short term SOM turnover in soils. In the future researchers should try to address spatial variability in more detail.
- In accordance with that we stated in our reply to referee #2, “our study aims at the estimation of gas fluxes (CO_2) from soil immediately after tillage to link them to short-term changes of SOM analysed by Py-FIMS”. Additionally, “the placement of our bases and their distances to one another were appropriate” since they integrate the high spatial variability of gas fluxes from soil (cf., Clough et al., 2015; Parkin and Venterea, 2010).
- The three collars and the distances between them had a circumference of at least 4.4 x 0.8 m in each treatment-plot. The soil samples were taken from outside of this area which means

that these samples integrated over an area of about 2.5 m², which may be on the edge, but was necessary to ensure the link to the gas fluxes.

- Additionally, we now refer to the master thesis by Jacobs (2014) who demonstrated that N_2O fluxes from the soil of the study site show very high small-scale variability well below the meter scale which tends to level off with increasing scale at the plot level: "Although the relatively small sampling areas around the bases in each treatment plot might suggest a 'pseudo-replication' in soil sampling, we have evidence suggesting a very high spatial variability in the soil, which alleviates this problem: In a master thesis on spatial variability, Jacobs (2014) revealed that N2O fluxes from the soil of the study site show very high small-scale variability well below the meter scale. Therefore, we assume 'real', i.e. independent replicates…" (Lines 466-470)

The observed differences were mostly quite small and just of few of them were statistically significant (particularly differences observed in the relative proportions of compound classes - table 2). Therefore, I recommend to adapt the whole discussion section taking the limitations of the design and the rather small differences into account.

- We acknowledge that the differences were often rather small. But concurrently, the detection of small differences is a specific feature and advantage of Py-FIMS. However, we added a section ("4.4 Limitations") concerning the observed small differences between the plots, which is also pointing at the reasonable detection of significant temporal changes: "though the comparison *between* the treatments should be done carefully because of possibly rather small differences. At the same time, due to the, thus, potentially lowered influence of spatial variability, our sampling design might have biased our results towards the detection of even small temporal changes *within* the treatments. Because we are mainly interested in the impact of tillage, this limitation is not interfering with our findings " (465-476)

The paper would really profit from shortening.

- You are right. The initial version of the MS was quite lengthy. However, already the now actual version on SOILD is substantially shorter due to edits in response of the comments of the two reviewers. In addition we carefully went through the MS again and tried to shorten it further, whereby also addressing your above raised point that we should adapt the whole discussion to the "limitations" of the study.

Some further comments:

Abstract:
Lines 19-20: I do not think that a higher proportion of volatilized matter during pyrolysis indicates an increased amount of SOM.

- You are right, additionally to higher amounts of SOM, increased proportions of volatised matter may also indicate a lower stability of SOM. However, the respective phrase was erased from the abstract in a later version of the manuscript. The context of volatised matter and SOM amount/stability is discussed in lines 385 ff.

Line 26: the high spatial variability is based on a personal communication – very weak base. I wouldn't use such an argument in the abstract.

- We agree that this phrase has no additional value for the abstract, therefore we removed it. However, the claim about the spatial variability is corroborated by own investigations as stated above.

Lines 27-28: increased in comparison to ???

- It increased in comparison to before tillage. This information is given in the previous sentence: "Significant changes in SOM composition were observed following tillage. In particular, … increased proportions of N-containing compounds in BD".

Material and methods:

Lines 91-92: The content of organic C in the soil is much higher than the reported ones in table 1. This difference was not considered in the discussion. Are there any reasons for that?

- Yes, there are. This was the amount reported from our previous paper. The sampling locations are ~50 m away from the spots we used here. We corrected this value to the averaged „Pre"-values from table 1 to not leave the reader with different numbers that cannot be understood without context. The values for bulk density and pH were obtained from the present spots before tillage.

I did not read any information about the history of the treatments. What kind of management has been applied before starting the experiment? Was the BD just once applied?

- Yes, BD was applied only once, because this trial was established for a single year. In the history, there were other trials at the site with winter wheat followed by maize. We edited the respective section to better reflect those details: "Before our study period, during other trials, winter wheat (*Triticum aestivum* L.) followed by maize were grown on the field. (102-103). "The BD for this single application originated from …" (112/113)

Lines 159-160: Please describe how samples were taken – composite samples of xy cores / subsamples??

- The samples were taken directly with the sample rings: "Three replicates of bulk soil samples were taken between 0 – 10 cm depth (depending on unevenness of soil surface due to tillage) directly with three soil sample rings ($h = 6.1$ cm, $V = 250$ cm^3)" (163-165) and were "… fixed immediately with liquid nitrogen and splitted thereafter into subsamples for freeze-drying and for oven-drying at 60° C." (163-170)

Lines 175-176: I do not understand the implications / functions of the averaged survey spectrum per treatment. Does it mean that the proportions of compound classes were calculated based on one spectrum per treatment? How did you calculate means and standard deviations?

- The averaged survey spectrum refers to the figure that we show. We redescribed the respective phrase: “For plotting, the three replicates of each sample were then averaged to one final survey spectrum” (181-182). For analysis we considered the replicate runs separately as replicates.

As one of the reviewers I did not understand the purpose of the hot water extraction.

The purpose of measuring HWC and HWN is now elucidated in the introduction after the outline about Py-FIMS: “Hot-water extraction is a relatively simple method to release labile SOM and to estimate how much of soil C and N can be easily utilised my microorganisms (Leinweber et al., 1995). These labile pools have been suggested to be an important indicator of short-term changes in SOM quality due to soil management (Haynes, 2005). Furthermore, a significant proportion of hot water-extracted organic matter originates from microbial biomass. Thus, this approach is a potential indicator for changes in microbial biomass or activity (Sparling et al., 1998), which may reflect sources of CO_2 efflux following tillage. (77-84)

Discussion

In the discussion I would prefer a more condensed way to discuss the most important findings of the study in a more straightforward way.

- This was acknowledged by the other referees, too, and, therefore, the recent MS should be already more concentrated. However, we have re-inspected the discussion based on your suggestions below and tried to tighten it even more.

I got the impression that OM in the digestate treatment was sometimes more and sometimes less available to microorganisms. That is not very consistent.

- Indeed there are two elements to the story of OM availability in the digestate treatment: On the one hand there is a generally increased availability of stable OM fractions from BD residues after tillage. On the other hand, there is a relative shortage of available C. For this reason, we refined the section on CO_2 which might have been confusing in this context: “… C originating from the digestates is likely less available to soil microorganisms compared to undigested organic matter …” (361-362). The discussion about Py-FIMS is following this differentiation quite consistently.

The discussion from lines 350-365 is not convincing. I would think that OM from BD might enhance soil aggregation, resulting in less disruption of aggregates during tillage.

- The lines you are referring to were already rearranged and mostly omitted in the most recent version of the MS, because the discussion about carbon use efficiency was too speculative. A short discussion on a higher resilience against disruption of macroaggregates due the use of an organic amendment has been introduced already: “On the other hand, even a single application of organic amendment can increase aggregate stability (Grandy et al., 2002).

Therefore, the resilience against disruption by tillage might be promoted, leading to a better physical protection of labile soil C not contained within digestates" (364-367).

Line 381-385: I do not see that, i.e. an enrichment of (stable) SOM indicated by a higher portion of VM.

- Here, we disagree. The share of VM as well as the total ion intensity are widely accepted as indicators for the amount and stability of SOM and SOC, respectively (Sorge et al., 1993; Wilcken et al., 1997; Leinweber and Schulten, 1995). The mean values of VM before tillage were 5.2% (BD), 3.9% (MF) and 3.6% (CL). To corroborate our position, we introduced also TII into this section. TII was 44.3 (BD), 34.2 (MF) and 41.5 (CL) x 10^6 counts mg^{-1}. Though we were not able to test for significant differences between the treatments, BD had the highest mean values for both variables. That's why we "suggest a tendency to elevated SOM" (387). Maybe the attribution of organic matter from BD application as "rather stable" causes confusing in this context. Thus, we omitted it and the respective section reads now: "VM as well as TII, which are indicators of SOM content (Sorge et al., 1993) and also of its stability (Ludwig et al., 2015), were larger in BD than in MF and CL before tillage (Table 2). This suggests a tendency to elevated SOM due to application of organic matter with biogas digestate." (385-387).

The discussion related to the rather small changes in hot water extractable N (and C) and relationships to Py-FIMS data is highly speculative and not very convincing. I would recommend to shorten this part substantially.

- We reconsidered this part and agree partly about the speculative character due to the rather small changes. We removed Figure 6 and most text bits about the single correlations between hot-water extracts and the compound classes in the results and the discussion sections, but we kept the correlations that were previously described by other authors, i.e., the relations between HWC and carbohydrates as well as between CO_2 and sterols. (312-323 and 435-444)

I would strengthen the discussion related to relations between CO2 effluxes and changes in SOM composition (starting from line 438).

- We followed your suggestion and strengthened this part. (445-452)

In the whole discussion about labile N I missed information about mineral N in the soil in the different treatments. This is the most available form of N.

- The signals of *m/z* 17 and 18 (NH_3 and NH_4) are discussed in the recent MS version (395-399) because we included it before in reaction to reviewer comments.

Tables and figures

The standard deviations as indicated in table 1 are quite low. That might be the result of the design of the study and does not support the argument of high spatial variability.

- In Table 1 that's true, but SD in Py-FIMS data (Table 2) are rather high and the latter method is much more sensitive to differences in compound classes than simple C and N determinations. However, after all, the MS is now focussed on turnover processes rather than on differences between fertilisation regimes.

Independently from your comments, we have added an additional reference supporting the idea that microorganisms utilise lignin-derived moieties at low initial availability of carbohydrates:
"Recently, Rinkes et al. (2016) also found that decomposers may break down lignin to acquire C for their metabolism in the absence of available labile C. (415-417)", added with a respective phrases in lines 453-460: "…enhanced microbial N-turnover by tillage in soils amended with biogas digestates; possible co-occurring with the decomposition of lignin as C source due to a relative shortage of carbohydrates. […] increase of N-containing compounds along with decomposition of lignins and formation of carbohydrates and peptides. "

References

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

Jacobs, O. (2014). Vergleich von zwei unterschiedlichen Kammersystemen zur Messung von bodenbürtigen Lachgasflüssen. Master thesis, University of Rostock

Leinweber, P. and Schulten, H. R.: Composition, stability and turnover of soil organic matter: investigations by off-line pyrolysis and direct pyrolysis-mass spectrometry, Journal of Analytical and Applied Pyrolysis, 32, 91–110, doi:10.1016/0165-2370(94)00832-L, 1995.

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.

Sorge, C., Müller, R., Leinweber, P., and Schulten, H.-R.: Pyrolysis-mass spectrometry of whole soils, soil particle-size fractions, litter materials and humic substances: statistical evaluation of sample weight, residue, volatilized matter and total ion intensity, Fresenius' Journal of Analytical Chemistry, 346, 697–703, doi:10.1007/BF00321275, 1993.

Wilcken, H., Sorge, C., and Schulten, H. R.: Molecular composition and chemometric differentiation and classification of soil organic matter in Podzol B-horizons, Geoderma, 76, 193–219, doi:10.1016/S0016-7061(96)00107-3, 1997.