Replies on referee comments on Van de Broek et al., The soil organic carbon stabilization potential of old and new wheat cultivars: a $^{13}CO2$ labelling study

Replies to Stefan Karlowsky

We would like to thank Dr. Karlowsky for his comprehensive and detailed comments on our manuscript. These will greatly improve the quality of our manuscript. We present the reviewer comments in *italic*, our replies are formulated in normal font.

General Comments

In the present study, the authors report their findings from a 13CO2 pulse labelling experiment on different wheat cultivars grown in lysimeters filled with agricultural soil (surface and subsoil). The main study objective is to assess how the use of more recent wheat cultivars with lower rooting depths and root biomass alters organic carbon inputs into soil compared to older cultivars from the Swiss wheat breeding program. This research subject is important, because a large share of the global agricultural land is allotted to the cultivation of cereal crops, and it is unclear how the use of mod ern cultivars with altered root traits affect soil organic carbon (SOC) dynamics and consequently SOC stabilisation. Here the authors found no significant effects of different wheat cultivars on SOC in the short term. They conclude that the fate of root biomass after the harvest determines cultivar effects on stabilised SOC pools in the long term. The study is based on a sophisticated methodological approach, including an innovative lysimeter-labelling chamber setup as well as state-of-the-art 13C labelling and analysis techniques. The description of materials and methods used for the study is, in general, detailed enough to follow all steps of the experiment. However, a few things still need clarification (see specific comments). The major limitations of the study are the low number of replicates (probably due to the complex setup) and the fact that root biomass was too low for 13C analysis in many samples. Especially the latter impedes drawing conclusions about the input of plant-derived carbon into soil and its variation between the different cultivars over the growing season. Notably, the authors are well aware of these limitations and discuss them appropriately.

The presentation of results is generally OK but should be modified in order to avoid redundancy between figures and tables. E.g. Fig. 1 and Table 1, both are showing the same values and statistics for aboveground biomass.

The data were presented in the figure for visual interpretation, while we repeated the values in the table so the exact values are available to the reader. In order to reduce further redundancy, we no longer cite these values in the text, but refer to the table.

Furthermore, you do not need to repeat values shown in figures and tables in the text body, neither in the results nor in the discussion section.

These values are now removed from the text.

I would also recommend to change Fig. 1 and Fig. 2, not separating between biomass and delta13C, instead showing aboveground biomass together with its delta 13C in Fig. 1 and the root parameters in Fig. 2 (as it is structured in the text body).

We thank the reviewer for this suggestion, but we think it's more convenient for the reader to see separate figures for OC % and $\delta^{13}C$, although this is discussed in the text differently. This way, for example, we aim to emphasize the important differences in the $\delta^{13}C$ value of above- and belowground biomass in the subsoil between the old and more recent cultivars.

Regarding Fig. 4 and Table 2, I am missing the statistics. These statistics would be necessary to support some of your interpretations from the discussion part.

We fully agree with the reviewer that statistics would aid our interpretation and discussion of the results. However, as explained at the end of section 2.2.4, due to (i) the large variability in root $\delta^{13}C$ among the different replicates and (ii) the low biomass of retrieved roots, which prevented $\delta^{13}C$ analyses for roots at certain depths for multiple lysimeters, we decided to calculate net rhizodeposition using the average values for all 3 replicates.

The variations on the final calculated values (as shown using error bars in Figure 4) was calculated by error propagation calculation, using the standard errors of values used for these calculations when they were available (i.e. when values were present for the three replicates of a certain cultivars, for example for SOC %, bulk density etc.). This is now explicitly stated in the caption of Figure 4: 'Error bars represent the standard error (n = 3), which was calculated as error propagation based on the standard errors for the average for the different cultivars. This prevented statistical analyses of significant differences between the cultivars.'.

Therefore, the variation within the cultivars was not be accounted for, which prevented us from performing statistical analyses here. We are aware that this limits our interpretations, but acknowledge this in the text (section 2.3.1), where the following sentence has now been added: 'Uncertainties on these calculations were assessed using error propagation of the variables for which standard errors could be calculated (i.e., for which values were available for the three replicates of a cultivar). When standard errors of the δ13C value of root biomass for a certain depth layer could not be calculated due to a low number of replicates, the standard error was calculated by multiplying the average relative standard error of the layers above and below this layer with the δ13C value of this layer.'.

The discussion itself is a bit lengthy and would profit from some restructuring (see also specific comments).

We used your comments below to shorten the discussion

The subsections 4.1 to 4.3 can be shortened, e.g. by excluding the repetition of results and streamlining the remaining text. Maybe it is also better to start the discussion section with the main study object (for non-expert readers), which suddenly comes up in subsection 4.4 now.

Thanks for this suggestion. We now start the discussion by briefly repeating the objective of our study and the main results, for readers who jump to the discussion section at once. Where possible, we shortened sections 4.1 to 4.3, e.g. by removing the first paragraph of section 4.2.

Another possibility to increase readability would be the use of more active and less passive voice, though this is a matter of taste. Overall, the structure of the manuscript is clear and the language is fine. The authors relate their work to a comprehensive set of up-to-date literature and make the data underlying the results available as supplementary material. However, there are a few things in need of improvement and the manuscript will profit from a revision taking into account the addressed points.

Specific Comments

Line25: I think that "net SOC stabilization" is the wrong term. Stabilization implies a long-term effect, which you did not study here (if there was a difference - what about rhizodeposits degraded and respired by microbes off-season?). Therefore, better use "net carbon rhizodeposition" as in the rest of the manuscript.

A similar comment was raised by the second reviewer as well, and we agree with both reviewers. Therefore, we changed the term 'carbon stabilization' to 'net rhizodeposition' throughout the text. We, however, did not change the title of the manuscript, as here we talk about 'carbon stabilization potential', and net rhizodeposition and root biomass (which we study) given an indication about the potential to stabilize carbon on the longer term.

Line 85: To my mind, this sentence is unnecessary, because the rationale of the study should be clear from the text above. I suggest starting directly with your research questions and marking them as such.

Thanks for this suggestion. We removed these sentences and explicitly formulated the research question and the hypothesis.

Line 143: Please indicate the approximate time of day when the labelling was carried out.

The labelling was carried out at 2 pm, this has now been added to the text.

Line 146: Was it always the same chamber/cultivar for monitoring CO2 concentrations?

The monitoring was always done at the same chamber and thus cultivar. The monitoring intended to approximate general CO2 uptake within for instance changing chamber volumes and less to adjust for each individual cultivar/chamber. We agree with the reviewer that this is not ideal but we had to consider technical implementations as well as time issues. Therefore it was decided to only monitor at one chamber. However, given the relatively similar enrichment across all cultivars in aboveground plant biomass we believe that the labeling was done relatively homogeneous. We added to the text that CO_2 concentrations were always measured in the same chamber: 'Throughout the experiment, CO2 concentrations were measured in the same chamber.'.

Line 149: Is there an estimate for the CO2 concentration at the end of the two hours?

No, the CO_2 concentration in the chambers was not measured after these two hours.

Line 158: What does "limited amount of samples" mean – only at the end of the experiment (i.e. data shown in Fig. 3)?

Yes, that is what we meant. We clarified this in the text: 'In addition, the $\delta^{13}C$ value of CO_2 was measured for CO_2 samples collected along the depth profiles on the last sampling date, using a Gasbench II […]'

Line 175: From my own experience, it is better to analyse soil microbial biomass directly from fresh (unfrozen) soil, because the freezing can increase the amount of carbon found in the non-fumigated fraction (probably cell lysis). However, regarding the delta13C values in comparison to SOC, this does not seem to be a problem here.

We are aware of the fact that this would indeed be a better practice. However, due to technical constraints we had to perform the measures on frozen soil samples.

Line 211: Did you use the same value of -28 ‰ for aboveground biomass?

We only had to make an assumption about the $\delta^{13}C$ value of roots, which was necessary to calculate the excess ^{13}C, to eventually calculate net rhizodeposition. As this was not done for aboveground biomass, we did not need to make assumptions about the $\delta^{13}C$ value of the unlabelled aboveground biomass.

Lines 217-218: This sentence is unclear. With "some of the input variables", do you mean biomass or delta13C?

This was related to the $\delta^{13}C$ of root biomass. This has now been added to the text: 'In addition, there was a large variability in the $\delta^{13}C$ value of root biomass between the replicates of the same cultivar, which complicated the calculation of excess ^{13}C for individual lysimeters.'

Line 225: Please explain why you used the Janzen and Bruisma's equation in addition to excess 13C. If I understand it correctly, Fig. 4A shows the summed values for all soil layers as excess 13C according to Eq. 3 and Fig. 4C shows the data for individual soil layers as rhizodeposition C according to Eq. 4. However, the unit in Fig. 4C (g m-2) rather points to excess 13C. This must be clarified.

That is correct: Eq. 3 was used to calculate the mass of recovered ^{13}C label (g ^{13}C m^{-2}), while Eq. 4 was used to calculate the total amount of net carbon rhizodeposition, using the excess ^{13}C in roots and the soil (g C m^{-2}). To make this more clear in Figure 4, the unit of the label of Fig 4a has been changed to (g ^{13}C m^{-2}), while the unit in the label if Fig 4c has been changed to (g C m^{-2}).

Line 284: Did you find a significant effect for the three blocks? Why did you use the blocks as fixed effects and not as random effects, i.e. error term, in the ANOVA? Please also report the significance levels for the different statistical tests. In general, I would prefer using the Tukey-HSD test, because it also accounts for multiple comparisons (in particular when depth is added as additional factor).

For some of the variables, we did find a significant effect of the blocks (e.g. aboveground biomass), while for other variables this was not the case (e.g. belowground biomass). For the analysis of statistical differences between properties of the cultivars (e.g. aboveground biomass), we used a two-way anova without interactions. This is generally recommended for the analysis of randomized complete block designs (e.g. Dean et al. (eds.), Handbook of design and analysis of experiments, ISBN 978-1-4665-0434-9, or https://stat.ethz.ch/~meier/teaching/anova/block-designs.html). Therefore, block was not treated as a random effect. We note that for the three-way anova, we included block as a random effect (see L. 289 – 291). The significance level for the Tukey's test is now added to this section: '[…] using a significance level of 0.05'.

Lines 296-300: The aboveground biomass values are repeatedly reported in the text, Fig. 1A and Table 1/Table S1. It is sufficient to show the results once, especially since all individual values are available in the supplementary excel file. Remove this redundancy.

Thanks for this suggestion. As stated above, we removed the values for aboveground biomass throughout the test. However, we prefer to show the values for aboveground biomass in Table 1, to give the reader a complete overview of the values of both above- and belowground biomass. We are aware of the fact that these values are shown in Figure 1, but we want the reader to be able to consult the exact values without having to go look for the online supplement.

Line 325: Interpretations/conclusions do not belong to the results section. Delete this sentence.

This sentence has been deleted.

Line 341: Note that the soil microbial biomass was higher in Zinal (Fig. S3), so that excess 13C was probably similar to Mont-Calme 268 (Fig. 4C).

Thanks a lot for this remark, we now included after that sentence: 'However, as the microbial biomass under Zinal was substantially higher compared to under Mont-Calme 268 in this layer, this does not necessarily imply that microbes under Mont-Calme 268 incorporated more excess ^{13}C compared to under Zinal'.

Line 357: Do you mean "statistically significant" with "substantially"? Unfortunately, no statistical information is provided in Fig. 4.

We meant substantially, since no statistical test could be performed (see above)

Lines 364-367: Please improve the sentence structure.

We changed these sentences to: 'The total amount of net carbon rhizodeposition measured at the end of the experiment down to 0.75 m decreased with depth for all wheat cultivars, with the exception of Zinal (Figure 4C). The highest values were observed for Probus (126 ± 57 g C m^{-2}), followed by CH Claro (112 ± 39 g C m^{-2}), Zinal (100 ± 39 g C m^{-2}) and Mont-Calme (85 ± 27 g C m^{-2}). There was thus no clear relationship between the amount of net carbon rhizodeposition and year of release of the wheat cultivars.'

Line 372: Do you have any explanation for the abrupt increase of Co2 concentrations?

We think this was caused by roots growing down to these depths at this moment, although we do not have conclusive evidence for this. For this reason, we do not elaborate on this in the manuscript.

Line 399: How are your results (no differences in root biomass between cultivars) in line the study of Friedli et al. (2019), showing substantially (statistically significant?) higher root biomass in older cultivars than in more recent ones? That is contradictory!

As stated in line 395 – 396, we did find differences in root biomass between old (161 & 205 g m^{-2}) and recent (107 & 97 g m^{-2}) wheat cultivars, although these were not statistically significantly different (due to large variations within cultivars). Friedli et al. (2019) found that cultivars from the Swiss wheat breeding program showed decreasing root biomass with increasing year of cultivar development. Therefore, we state that our results are 'in line' with the results from Friedli et al. However, we have emphasized that are results are not statistically different (L 397 - 399).

Line 406: To which species does the root:shoot ratio of 0.14 belongs to, is it an average value?

This indeed is the average value for all the cultivars studied by Friedli et al.. This has been clarified in the text: '[…] including the cultivars used in our study (an average value of 0.14 for all cultivars studied by Friedli et al. (2019)).'

Lines 429-444: This paragraph reads like an introduction passage. It is better to move it to delete it from the discussion and combine it with overlapping parts of the introduction.

This comment was also raised by the other reviewer and we agree that this paragraph is redundant here. Therefore, we deleted this paragraph to reduce the length of the discussion, as part of this is covered in the introduction.

Lines 459-471: The repetition of results should be avoided and the two paragraphs streamlined to 2-3 short sentences.

Thanks for this suggestion, these 2 paragraphs can indeed be shortened considerably. We chose to retain the values we provide about the total amount of carbon that is allocated belowground, as this is not reported elsewhere in the manuscript, so we can compare them to literature values.

Line 492: By "assess the effect of wheat cultivars from a century of wheat breeding", do you mean that you assessed the effect of four cultivars representative for changes during a century of wheat breeding?

Yes, that is indeed what we meant. As we now re-stated the aim of our study at the beginning of the discussions (see above), we removed this sentence here, as it is redundant.

Line 494: There is no statistical support for this statement, neither for root biomass nor for belowground carbon allocation. In consequence, it is not surprising that you did not find effects on the SOC pool according to the next sentence.

See about the inability to statistically prove this in previous responses. We did find that there were no statistically significant differences between the root biomass of different cultivars (Table 1), although the averages suggest that root biomass was larger for the older cultivars. To clarify this, we included '[…] allocated more assimilated carbon belowground, although this could not be statistically proven'.

Lines 506-509: This cannot be generalized, because the activity and substrate preference of microbial communities depends on a variety of factors (e.g. Delgado-Baquerizo et al., 2016:

https://doi.org/10.1111/1462-2920.13642). In addition, the preference for recent plant-derived substrates or more stable SOM varies with soil depth (Kramer & Gleixner, 2008: https://doi.org/10.1016/j.soilbio.2007.09.016) and the presence/quality of plant residues is known to alter soil microbial communities (e.g. Bai et al, 2016: http://dx.doi.org/10.1016/j.apsoil.2015.09.009). In this sense, the microbial community can be shifted to more fungi and Gram-positive bacteria in the presence of more complex organic compounds derived from root residues.

We agree with the reviewer that the fate of roots in the subsoil (mineralisation versus stabilization) is more complex than as we stated in the manuscript. Therefore, we shortened this section, as this is not the focus of our study, while briefly also incorporating the remarks raised by the reviewer: 'However, it is not straightforward to make predictions about the amount of root biomass that will be stabilized in the soil in the long term, as this depends on the efficiency with which plant-derived biomass is incorporated in microbial biomass (Cotrufo et al., 2013) and interactions between soil depth, the microbial community composition and its substrate preference (e.g. Kramer and Gleixner, 2008), among other factors.'.

Line 519: There is no statistically significant difference, only a slight trend.

We changed this sentence to: 'In contrast, despite the lack of statistical evidence, we observed differences […]'.

Technical Corrections

Lines 47-49: Please reformulate the two passages with "is also proposed". This is very repetitive, since the term "has been proposed" is already present in Lines 45-46.

Thanks for noticing this, we removed 2 of the 3 'is proposed' by an alternative wording.

Line 146: Obviously, this should be 40 g and not 40 mg.

I assume you meant line 176? Here, is should indeed be 40 g, thanks for noticing this.

Line 179: This sentence fits better in the previous subsection at line 172.

This sentence is at this location because the determination of the gravimetric moisture content was necessary to calculate microbial biomass carbon per unit dry soil. To make this clear to the reader, this sentence was changed to: 'To determine microbial biomass carbon per unit of dry soil, the gravimetric soil water content was determined by drying about 10 g of each soil sample at 105 °C and subtracting the weights before and after drying.'. We note that we also determined the gravimetric soil moisture content for bulk soil samples collected from the lysimeters at the end of the experiment. This is mentioned in line 171.

Line 193: Technically, you measured the delta 13C of C-F and C-NF instead of microbial biomass.

Thanks for pointing this out, we changed this in the manuscript: '[…] fumigated and non-fumigated soil (for the determination of microbial biomass C and $\delta^{13}C$) […]'.

Lines 249-250: Separate the "s" and "i-1,i"/"i,i+1" in the formulas (maybe by a semicolon), as it can be confusing otherwise.

Thanks for this suggestion, we changed this accordingly.

Line 350: Replace "showed substantial variation" with "varied".

This has been changed

Line 381: Include "biomass" (Plant biomass, carbon dynamics…).

This has been changed

Line 416 "were respiring CO2 down to greater depths…" -> Reformulate.

This was reformulated to '[…] roots of the old wheat cultivars respired CO_2 at greater depths compared to […]'.

Line 453: The shown references do not include only the same studies.

Thanks for pointing this out. We now shortened and combined both sentences, without mentioning the 'same studies'.

Line 485: Twice "assess/ed"

Thanks, this has been replaced