Authors response to Reviewer 1 (RC 1) for soil-2021-81

In their manuscript, 'Soil organic matter and labile fractions depend on specific local parameter combinations', Ortner et al. present their work on the analysis of factors controlling the soil organic carbon (SOC) concentration in topsoils of the region around Trier, Germany. The authors collected topsoil samples in arable land and grassland in 4 regions with different parent material, and determined the organic carbon (OC) concentration, hot water extractable carbon (HWEC), microbial biomass carbon (MBC) and multiple soil properties on these samples. They used PCA to cluster the soil samples based on parent material and soil texture into different clusters. The aim of their study was to assess the main factors controlling topsoil organic carbon concentration, HWEC and MBC using two modelling approaches: a bivariate model and mixed effects models. The main findings are that (i) mixed effect models outperformed bivariate (linear) models in predicting OC%, HWEC and MBC, (ii) at the local scale, site-specific parameters explained OC variability better than landscape-related variables and (iii) using the 'local' model resulted in better results when predicting the OC% of a specific cluster compared to the 'global' model.

The results of the present study help to improve our understanding of the factors controlling topsoil organic carbon concentrations at the landscape scale, which is needed e.g. in order to improve soil organic carbon models. The authors have constructed a valuable dataset which may benefit other researchers. I would therefore encourage the authors to make this data available through an online repository, instead of making it only available upon request.

We thank for this comment. The dataset was generated in the framework of a contract project of the UBA. We aim to clarify with UBA if the data can be fully published.

Overall, the manuscript is well-written. However, at multiple locations very long sentences are used, which does not benefit a smooth reading. Splitting those sentences and using more commas would improve the readability of the manuscript considerably. In addition, I would encourage the authors to use subsections in the Results and Discussion sections, which will provide a better overview to the reader of what is being presented and discussed.

Following the advice, we revised and split several of the longer sentences. Further, we took up the valuable hints regarding subsections for the results and discussion sections.

One of my main concerns about the present manuscript is related to the quantification of the goodness-of-fit of the different models, which is now done using R-square. This is a measure to quantify the proportion of variation in a dependent variable that is explained by an independent variable, but is not a measure for the goodness-of-fit of a model. For example, a very poor model can have a high R-square value, while a good model can have a relatively low R-square value. Therefore, the authors should use a different measure to quantify the goodness-of-fit of their model when comparing measured with modelled data, such as the (root) mean square error or similar.
We agree that R2 is good to show the percentage of explained variance but not fully sufficient to document the goodness-of-fit of a multivariate and/or non-linear

model. Hence, the RMSE was added as a measure for the goodness-of-fit. The presentation of R2 was reduced to the bivariate linear models.

In addition, I missed a discussion about the broader implications of the results and the implications for future research. For example, do the authors suggest that researchers should use 'local' model whenever possible? And how about regions where local information is not present? It would also be very informative if the authors would quantify the difference in predicted SOC% when using a global versus local model. To how much of an over or underestimation would this lead? Is that difference significant enough to invest more resources in the collection of local data?
Statements about broader implications and some recommendations for further research were added to the conclusions.
The level of over- or underestimation is represented by the RMSE. We added the RMSE

Another concern is related to the title, which I find not very informative. For example, it will not be clear to someone who has not read the manuscript what 'specific local parameter combinations' are. Also, it would be good to be more specific about what they mean with 'soil organic matter and labile fractions'. From the title it is not clear if the authors mean SOC concentrations, stock, spatial distribution etc. In addition, the manuscript discusses soil organic carbon, and not all soil organic matter, so would be good to change this in the title.
We changed the title in order to make it clearer: "Content of soil organic carbon and labile fractions depend on local combinations of mineral phase parameter"

Lastly, it would good if the authors specify in the beginning of the manuscript that they discuss SOC concentrations, and not stocks. Throughout the manuscript, the authors talk about 'SOC' without specifying that it concerns concentrations, not stocks. This is an important difference, which should be clear to the reader from the abstract onwards, and repeated throughout the manuscript. For example, the authors could change 'SOC' to 'SOC%' to make this clear.
In the text (e.g. Abstract) and title it is now explicitly mentioned that concentrations were investigated.

Answers to specific comments RC1

L18: Would be good to explain here what you mean with 'global' and 'local' clusters (and models).
Thank you for this hint, we added the definition for the investigated clusters. Further we decided to replace the term 'global' by 'total' to prevent any confusion regarding different scales (local vs. global scale). It should be now clearer that we talk about the total dataset encompassing the different local datasets.

L19: define that you assess SOC concentrations, and thus not stocks
See general comments. We added this information to the title, Abstract as well as in the Introduction and in Material and Methods. In Tables and Figures SOC is given in % as unit, indicating that we assessed concentrations.

L20: would be good to explain here which 'labile fractions' you study
Information regarding the labile fractions was added as requested. "… explaining SOC and its labile fractions hot water-extractable C (HWEC) and microbial biomass C (MBC)".

L27: here you use the term 'organic matter', while until here you used 'SOC'. Please be consistent with these terms, and only use one
It was changed to 'organic carbon'.

L29-30: 'showing that labile fractions depend less on soil properties than on organic matter input and turnover in soil'. The latter were not studied, so you cannot say this with certainty. Would be better to end the abstract with a statement about the broader implications of your results.
We thank for that comment. To avoid the impression that organic matter input and turnover were investigated in this study, we changed the sentence: "showing that labile fractions depend less on soil properties but presumably more on processes such as organic carbon input and turnover in soil."

L41: another important labile fraction of SOC is particulate organic carbon. Would be good to justify why you did not study this fraction
The authors fully agree that particulate organic carbon (POC) is an important labile fraction. Due to trivial financial reasons we had to decide which fraction(s) we can study. We decided for hot water extractable carbon (HWEC) and microbial biomass carbon (MBC) because they are methodically clearly defined. Quite often it is stated in the literature that both are very closely correlated with each other, and thus deliver no different information. We hypothesized and aimed to show that this is not the case (which was confirmed in this study). Additionally, we decided against POC because it is not uniformly defined, either by size or by density. So we hope that HWEC and MBC are representative measures of labile SOC pools. Again, we fully agree that having additional data on POC would have been a perfect completion of the dataset.

L45: 'MBC is expedient to explain SOC dynamics': this is rather vague, please be more specific
To make the point more clear, we added the following sentence: "Additionally, labile carbon fractions such as MBC quantitatively dominate in short-term turnover processes, while changes in SOC will only become significant over periods of decades. Therefore, MBC is expedient to explain SOC dynamics".

L45-46: 'much less research and attempts for quantitative modelling of these labile fractions […]': recently, multiple mechanistic models have been used to simulate labile carbon fraction such as MBC and POC, e.g. Ahrens et al. (2015), Wieder et al. (2015) and Zhang et al. (2021)
Thanks for this valuable comment. We changed the sentence and included some recent publications on modelling. We left the statement that SOC is mostly considered for such simulations, while there is still a need to take labile fractions more into account in order to gain a better understanding of SOC dynamics.

L58: 'In addition or even instead of': choose one
Ok, done.

L63: please clarify what you mean with parameters versus factors, as you use these terms throughout the manuscript
We added some examples. In general parameters include soil properties on a interval or ratio level of measurement while factors were applied on a nominal level of measurement.

L87: please define what you mean with 'global models'

To avoid confusion, we replaced both terms 'global dataset' and 'global model' with 'total dataset' and 'total model'. The total model is based on all data of the total dataset that encompasses all local datasets.

L99: the term 'sufficient quantification' is rather vague, please clarify this
The sentence was changed as follows: "It was aimed to determine the suitability of local models in comparison to total models to achieve an improved quantification of SOC, HWEC and MBC for local landscapes with distinct properties.".

L107: 'similar numbers of samples': how many per region?
Number of samples taken per region were shown in Table 1. Additionally, they were now added to the sentence in brackets for each sampling region.

L108-109: the use of the abbreviations throughout the manuscript is not intuitive and confusing for the reader, please use different names to identify the different regions, e.g. the parent material
We agree that abbreviations are a compromise between clarity and readability. Using the full terms or terms such as 'Muschelkalk' and 'Luxemburg sandstone' would have been too long, though. Shorter abbreviations were also inconclusive, e.g. schist and sandstone are both abbreviated 'S'. Hence, we plead for keeping the chosen abbreviations.

L119: why were some samples stored at -20 °C and others air-dried?
Samples were stored until they were analyzed. Storage was done in a uniform way for all samples. One part of each sample was air dried for subsequent chemical and physical soil analysis, another part was kept moist and was frozen for subsequent soil microbial analyses (MBC, MBN or respiration). This is now clarified in the text.

L134: was the chloroform fumigation extraction performed on samples freshly collected from the field?
Chloroform fumigation extraction was done on sieved material that was stored at -20°C before analysis. This was done to avoid changes until measurement was conducted. The suitability of this storage was proven in preliminary projects (data not shown).

L137-140: for how long were the samples incubated? How often was the CO2 measured?
Samples were preincubated at room temperature for one week (7 days), measurement was conducted for 24 hours at an interval of one hour. The information was added to the text.

L143: were all parameters log-transformed? Please clarify this
To conduct the principal component analysis all variables were log transformed to receive standardized and comparable variables. The information is contained in the text.

L146: please provide some examples of the 'mineral phase parameters'
We added two examples (Fe_o and fSilt+clay) into this sentence.

L146-147: please provide more information about the linear regressions that were performed

We added the information that we applied linear regressions using single predictors, and information that we checked the residuals for normality.

L156-157: Please provide information about which parameters were removed from the models
The non-significant parameter with the highest p-value was removed from the model. This was repeated until all remaining parameters were significantly contributing to SOC, HWEC or MBC. This information was added to the sentence.

L161: were all parameters checked for collinearity? Please clarify
We checked all mineral phase parameters for collinearity, which were used by the mixed effect models. Based on this, it was found that soil texture components (Sand, c+mSilt and fSilt+clay) showed collinearity as well as ECEC and $Ca+Mg_{ECEC.}$ We clarified this in the text.

L163: why a square root transformation? Please justify this
Square root transformation was selected as a common transformation and was suited to achieve normal distribution and heteroscedasticity of the residuals.

L163: Please clarify how the performance of the models was examined
Basically, we started by comparing the explained variance and, based on your valuable comments, now also added RMSE as indicator for performance.

L170: Please clarify the difference between 'soil' and 'topsoil' properties
Topsoil was separately mentioned due to the fact that our study focusses on agricultural topsoils. To avoid confusion or misunderstanding we decided to use only the term 'Soil properties'.

L177: are those differences statistically significant? What are the averages for the different parent materials?
There are some statistically significant differences, averages for the parent materials are given in Table 1 as mean ± sd.

L186: please provide examples for the 'parameters describing the composition of SOM'
We now mention some examples in the text, such as SOC, Nitrogen, hydrogen or oxygen, HWEC or MBC.

L191-194: this is not clear
We rephrased these sentences to make it clearer. Clusters identified by the PCA cover a different number of samples of the total dataset. Based on this clusters including the vast majority of samples were considered to represent the total dataset, while substantially distinct clusters, including only a part of all samples, were considered to represent local datasets.

L205-206: which 10 parameters?
Selected parameters were shown in Table 1 and in Table 3. Further we mention examples of these parameters in the Material and methods section. Examples of these parameters were added to the text in brackets.

L213: 'that largely matched with those found for the complete dataset': this is not clear
We adapted this sentence to make it clearer.

L224: what do you mean with 'sufficient extent'? Similar wording is used throughout the manuscript, but this is very subjective and should be clarified.
Thanks for this hint, we checked the manuscript and exchanged such phrasings by objective formulations using statistical parameters is applicable. See also L369

L240: please clarify what 'equal weight of samples' means
It means that both clusters (arable and grassland) contain a similar number of samples from each parent material resulting in a broad range for each soil property, catching up the properties from soils of each sampling region. We rephrased the sentence to clarify its meaning.

L237-242: please make clear that you are discussing the results of the bivariate models
We now mention that these lines address the bivariate regressions.

L241: what are the 'complex interactions of several different parameters'?
The term 'complex' was deleted. It makes sense concerning the environmental interaction of these parameters but not concerning the contribution to a mathematical model.

L243: please clarify what you mean with 'insufficient'. Which measure do you use to determine if a model performance is sufficient or not?
The authors thank for this hint, we changed such phrasings to objective formulations.

L249-251: R-square values are no measure for model performance, please provide the root mean square error (or a similar measure). Please show these results in a graph, perhaps in the supplement?
R-square is used to show the explained variance, this manuscript aims to show how much mineral phase parameters and their different combinations are able to explain the variance of SOC, HWEC and MBC. Notwithstanding, we fully agree that the root mean square error is a much better measure to determine the model performance. Therefore we added it to the text.

L273: do you mean the bivariate models with 'linear regressions' Please be consistent with this terminology
Yes, it means the bivariate models, we made it clearer.

L284: please replace R-square with a measure of model performance
See the above response. Further, as a measure for mixed effect models we added marginal and conditional R^2 to describe the R^2 directly related to these models.
R^2 based on predictions is only able to give a pseudo R^2 which is based on a linear regression between predicted vs measured. Such comparison between predicted vs measured and the received pseudo R^2 was technically the only option to test the performance of a total model, when applied on a local dataset. This information was added to the text.

L287: please provide the goodness-of-fit values before concluding that a certain model has an 'inferior performance'
We added this information, but we also kept the R^2 because it was aimed to investigate which model explained the variance to the highest extent.

L309-310: by saying 'Al- and Fe-oxides were shown to have a relevant influence on sequestration and stabilization of SOC', it seems like you explicitly studied this, while you only used a statistical model to assess this. Also, since you model SOC concentrations and not stocks, you cannot say anything about C sequestration, as this also depends on bulk density.
This sentence was linked to a reference and started with the term 'accordingly' in order to emphasize that this mechanistic interpretation of our statistical finding is not based on our study. We deleted the term 'sequestration' as requested since we do not address SOC stocks.

L342: please explain what you mean by 'multidimensional'
Multidimensional means that SOC is simultaneously affected by serval soil properties and factors which explain the overall accumulation and variability instead of single one to one interactions.

L348: 'to explain SOC': please clarify which aspect of SOC
The factors mentioned in the sentence were able to explain SOC under different scales and environmental conditions but in general, these factors enable to explain the accumulation and the variability of SOC.

L366: how do you conclude that sample size biased the results? Did you test for this?
It is true, randomly selecting a data subset from a total dataset does not necessarily lead to different (biased) modelling results. However, in this study total clusters including a larger number of samples showed a higher explained variance, which is a consequence of a broader variety of soil properties in the dataset. Local clusters with a smaller sample size also showed smaller ranges of the tested soil properties, leading to models with a lower explained variance.

L369: 'satisfying extent': how is it quantified that a model performs satisfying? Please be objective in deciding if a model is good or not
Thanks for this hint, we checked the manuscript and exchanged such phrasings by objective wordings.

L370: what do you mean with 'partially practicable'?
We replaced the term with 'less suitable', which is based on the lower statistical performance.

L374-375: 'sufficient quality level': same remark as L369
Similar to comment to L369, we changed the wording.

L380-381: by saying 'It became clear that […] with different annual dynamics', it seems like you tested annual dynamics. Please rephrase
We rephrased this sentence accordingly.

L281-382: You did not take SOC turnover into account, so how can you say that this is a reason for the lower explained variability by the different models?
An aim of this study was to investigate the linkage between mineral phase properties and labile fractions. Compared to SOC we found a lower explained variance for the labile fractions. Hence, we assume that – although we didn't explicitly investigate it - the known faster turnover of these fractions (depending, e.g. on land use management) will significantly contribute to the concentration of HWEC and SOC, thus explaining the gap in explained variance of HWEC and MBC.

L403-404: ‘sufficient estimation’: same remark as L369
Similar to comment to L369, we exchanged this formulation.

L405: would be good to end the Conclusions section with a statement about the broaderimplications of your results
We added the following statement: “Our research shows that local models, respecting site-specific parameter combinations, are superior to total models, although they are based on much smaller datasets. If available, they should be preferred.”

Figures and tables

Fig. 2: the colours in B are difficult to distinguish
It was changed accordingly.

Table 3: please make clear in the caption that these are the result for the bivariate Regressions
It was changed accordingly.
Fig. 3: ‘Predicted vs. measured’: please clarify in the caption which model was used to make these predictions. Please provide a measure for the goodness of fit and remove the R-square values, as this is not measure for model performance
We now mention it in the caption and added RMSE as measure for model performance.

Table 5: please provide more information about the table in the caption, the table should be clear to the reader without having read the entire manuscript. It would be more informative to provide a table with e.g. root mean square errors instead of R-square
We added some information regarding the RMSE, but we also want to show how the models differ in their explained variance. So we kept R^2.

Figure 6: it would be informative to show the same graphs for other clusters in the supplement. Please remove the R-square values and replace them by a measure for the goodness-of-fit of the models
Fig. 6 shows the performance of the previously developed total model, when applied to a local dataset. The model was not fitted to the data of the local dataset (which would have yielded the local model). Consequently, pseudo R^2 is given as a measure to compare the agreement (or disagreement) of modelled vs. measured data. Additional, we also added the RMSE to this Figure.

Technical comments

L36: driver => drivers
Done

L57: expedient => suitable
Changed

L73: space between ‘asCa2+’
Changed

L119: it’s not clear what ‘respectively’ refers to
Removed due to changes in this sentence.

L170: it's not clear what 'respectively' refers to
Ok, rephrased

L172: it's not clear what 'thereby' and 'essentially' refer to
Ok, rephrased

L252: it's not clear what 'respectively' refers to
Ok, rephrased

L275: what do you mean with 'not last'?
We want to highlight that soil acidity and its describing parameter were also relevant. The typo, however, was corrected to 'not least'.
L304: 'the in total very sandy soils': please rephrase
Ok, rephrased
L320: what is 'circumneutral'?
Circumneutral means soil pH that is close to neutral or neutral having a pH between 6.5 and 7.5. It is an established term. See for example:
Carl O. Moses, Janet S. Herman, 1991, Pyrite oxidation at circumneutral pH, Geochimica et Cosmochimica Acta 55/2, 471-482.

L324: please remove 'respectively'
Done.

L343: please remove 'respectively'
Done.
L344: confirmed => is in line with
Done.

References

These references were chosen based on their scientific content. I leave it up to the authors to decide if they wish to include them in their manuscript.
Thanks for this valuable references, we added some of them to our manuscript.

Ahrens, B., Braakhekke, M.C., Guggenberger, G., Schrumpf, M., Reichstein, M., 2015.Contribution of sorption, DOC transport and microbial interactions to the 14C age of a soil organic carbon profile: Insights from a calibrated process model. Soil Biology and Biochemistry 88, 390–402. https://doi.org/10.1016/j.soilbio.2015.06.008

Wieder, W.R., Grandy, A.S., Kallenbach, C.M., Taylor, P.G., Bonan, G.B., 2015. Representing life in the Earth system with soil microbial functional traits in the MIMICS model. Geoscientific Model Development 8, 1789–1808. https://doi.org/10.5194/gmd-8-1789-2015

Zhang, Y., Lavallee, J.M., Robertson, A.D., Even, R., Ogle, S.M., Paustian, K., Cotrufo, M.F., 2021. Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model. Biogeosciences 18, 3147–3171. https://doi.org/10.5194/bg-18-3147-2021