Dear editors and reviewers,

we, the authors, thank you very much for the revision of the manuscript. Subsequently, we describe a point-by-point response to the reviews, followed by a list of all changes made in the manuscript.

CEC1: 'Comment on egusphere-2023-760', Juan Antonio Añel, 31 Jul 2023

Unfortunately, after checking your manuscript, it has come to our attention that it does not comply with our "Code and Data Policy".

https://www.geoscientific-model-development.net/policies/code_and_data_policy.html

Firstly, you have archived the MONICA code on GitHub. However, GitHub is not a suitable repository for scientific publication. GitHub itself instructs authors to use other alternatives for long-term archival and publishing, such as Zenodo. Therefore, please, publish your code in one of the appropriate repositories, and reply to this comment with the relevant information (link and DOI) as soon as possible, as it should be available for the Discussions stage.

AC1: 'Reply on CEC1', Konstantin Aiteew, 03 Aug 2023

Since I'm not a developer of the MONICA model, I have no rights to the intellectual property of the model. My contribution consisted only of minor modifications to the code. In consultation with my colleagues, we decided to only make the entire source code available to the reviewers to avoid any legal difficulties. Therefore, as suggested I published the model code under a restricted access with Zenodo.

https://doi.org/10.5281/zenodo.8211378

CEC1: 'Comment on egusphere-2023-760', Juan Antonio Añel, 31 Jul 2023

Please, note that you should not include in your manuscript a "Software availability" section. The section for the information that you include there is the "Code availability" section.

AC1: 'Reply on CEC1', Konstantin Aiteew, 03 Aug 2023

The source code without the modifications is published by the actual developers and I used the section "Software availability" to describe the original model and the project website. In the revised manuscript I will rewrite this point and include it to the "Code availiability" section. I still recommend to visit the model website https://github.com/zalf-rpm/monica for model description.

CEC1: 'Comment on egusphere-2023-760', Juan Antonio Añel, 31 Jul 2023

Secondly, it is not enough that you state in the "Data availability" section that it is possible to obtain the input data from the "State Office for Mining, Energy and Geology of the Lower Saxony state government". This information is hardly useful for anyone that wants to get access to the data. You should include full details about the webpage, email addresses, persons or processes that anyone needs to follow to obtain them. Ideally, you should store a copy of the data that you have used in a private Zenodo repository. In this way, the dataset continues to be private, and we are sure that it is properly stored, and you would have a DOI to include in your manuscript and cite.

AC1: 'Reply on CEC1', Konstantin Aiteew, 03 Aug 2023

As suggested, I did the same with the data. The whole dataset is published under restricted access with Zenodo.

https://doi.org/10.5281/zenodo.8211468

I will give more information about the data used in the "Data availability" section. However, as far as I know, obtaining of the long-term monitoring data from Lower-Saxony, Germany is only possible if you work for the Federal Republic of Germany.

CEC1: 'Comment on egusphere-2023-760', Juan Antonio Añel, 31 Jul 2023

Also, you say, "Additional data is freely available under the CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE"; however, you do not provide details on the data you refer to. You must clarify it and be sure that such data complies with the policies of our journal.

AC1: 'Reply on CEC1', Konstantin Aiteew, 03 Aug 2023

The described long-term soil monitoring data is available from the State Office for Mining, Energy and Geology of the Lower Saxony state government. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the State Office for Mining, Energy and Geology of the Lower Saxony state government from the website: https://www.lbeg.niedersachsen.de/boden_grundwasser/bodenmonitoring/bodendauerbeobachtung/das-boden-dauerbeobachtungsprogramm-von-niedersachsen-572.html and mail: bodenkundlicheberatung@lbeg.niedersachsen.de Additional data for evaluating soil temperatures and moisture contents is available from the Thünen Institute of Climate-Smart Agriculture from the website: https://www.thuenen.de/en/institutes/climate-smart-agriculture and mail: ak@thuenen.de, from the University of Hohenheim from the website: https://spurengas.uni-hohenheim.de/en/108520 and mail: ruser@uni-hohenheim.de.

CEC2: 'Reply on AC1', Juan Antonio Añel, 03 Aug 2023

MONICA, as stated in its GitHub repository, is published under the MPL-2.0 license. This means that you have the right to copy and redistribute the code. Therefore, the ZENODO repository must be public and include both the original source code and the modified or added code you have contributed.

AC5: 'Reply on CEC2', Konstantin Aiteew, 26 Sep 2023

We have also included both the original and modified source code of the MONICA model as open access.

The original and modified MONICA model source code is stored in the following repository.

https://doi.org/10.5281/zenodo.8380341

CEC2: 'Reply on AC1', Juan Antonio Añel, 03 Aug 2023

We can understand the problems with the data from the Lower Saxony state government. However, I recommend you contact them and ask to release the dataset open to everyone.

AC5: 'Reply on CEC2', Konstantin Aiteew, 26 Sep 2023

With the exception of the data from the Lower Saxony state government, as they contain sensitive management information from farmers. The long-term soil monitoring data from the Lower Saxony state government is stored in the following repository. However, it is restricted to the reviewers of the corresponding manuscript.

https://doi.org/10.5281/zenodo.8379057

CEC2: 'Reply on AC1', Juan Antonio Añel, 03 Aug 2023

However, we can not accept the restrictions for the data from the Thünen Institute and the University of Hohenheim. The first one is your institution, so what precludes you from publishing the data is unclear. The second one is a university; again, it is unclear what legal reason forbids a university from publishing data.

AC5: 'Reply on CEC2', Konstantin Aiteew, 26 Sep 2023

we have managed to come in agreement with the stakeholders and are able to release the datasets open to everyone.

The openly available data is stored in the following repository.

https://doi.org/10.5281/zenodo.8380332

RC1: 'Comment on egusphere-2023-760', Lorenzo Menichetti, 01 Aug 2023

Your priors are thin, though, even if uniform. A broader prior choice would probably yield different results, particularly regarding the model biases you found (like overestimating water content, for example) and model performance results. Your prior choice influences the results quite heavily. Do you have anything to add in this respect?

AC2: 'Reply on RC1', Konstantin Aiteew, 04 Aug 2023

We selected the prior values based on the observed upper and lower bounds of each parameter to ensure that each calibrated value is within the realistic, observed range. It is highly probable, that with a different prior probability distribution, deviating values could be achieved. However, it was particularly important to us that the parameter values were based on realistic and observed values. Based on that, we could go more into detail explaining the priors in the manuscript. Aside from that, when considering the performance of the model, we would not expect that the calibration could significantly improve the model performance. For comparison, we also applied a simple hill-climbing algorithm to estimate optimal parameter values and didn't achieve a better performance. We would also argue that complex biogeochemical models tend to perform weaker compared to simple C-turnover models and we were surprised that the MONICA model performed comparably to the simple models.

If considering the overestimation in the soil water submodel, we didn't evaluate the feedback between plant growth and soil water contents, but we assume that there should be at least a minor effect of the plant growth parametrisation on soil water changes. We don't think that the parametrisation of the C-turnover submodel had a significant effect on modelled soil water. We also used the default

parametrisation of the soil water submodel, so there could be a general bias in the parametrisation. However, this part needs further evaluation and we are not sure if this should be evaluated/discussed in this manuscript.

RC1: 'Comment on egusphere-2023-760', Lorenzo Menichetti, 01 Aug 2023

Compared to the simpler models, I expect better absolute performances (considering the best sets) of MONICA with broader priors but a lot higher indetermination and uncertainty in predictions.

Nothing wrong here, but as a suggestion, you could develop more as a central question the potential increase in indetermination of a more complex model (with interactions, which seem nonlinear) like this one compared to simple decomposition models. Adding nonlinearities to consider feedback (if I understand the model correctly) is the main potential issue and advantage of going more complex on top of data requirements.

AC2: 'Reply on RC1', Konstantin Aiteew, 04 Aug 2023

We also like the suggestion, to further clarify the difficulties of complex models, that they are inherently black boxes that make understanding the various factors/feedbacks difficult. Therefore, we used a sensitivity analysis, to estimate the effect of plant growth (if that is what you mean with fertility feedback mechanisms) and other factors on soil carbon changes. But we agree that this aspect could be developed further in the manuscript.

RC1: 'Comment on egusphere-2023-760', Lorenzo Menichetti, 01 Aug 2023

Line 155-160: so you are using linear allometric ratios with no intercept, right? This could deserve some discussion if you like (the approach works well within certain limits, which are likely including your test sites anyway since you don't seem to have anything extreme)

AC2: 'Reply on RC1', Konstantin Aiteew, 04 Aug 2023

Line 155-160: We used the allometric functions to estimate the C-Inputs from the harvested yields, on the one hand for the simple C-models and second for comparison with the modelled C-Inputs of the MONICA model.

RC1: 'Comment on egusphere-2023-760', Lorenzo Menichetti, 01 Aug 2023

Line 232: define better the squared range between 1% and 5%. Does this range vary with different parameters?

AC2: 'Reply on RC1', Konstantin Aiteew, 04 Aug 2023

Line 232: This range seems quite arbitrary, but it was used in another publication and it worked quite well to achieve a random walk in the parameter space. Basically, each step for generating new candidate values was limited by 1 – 25 % of the parameter space. We assume the formulation we used was to complicated. We try to explain this aspect a bit better in the revised version.

RC2: 'Reply on AC2', Lorenzo Menichetti, 07 Aug 2023

I am surprised you expect a more complex model to perform worse than a simpler one. Based on that, I would expect the opposite; the more complex a model is, the more freedom it has to fit the information

content of your calibration data. I often borrow the bias-variance trade-off concept from machine learning, at least conceptually, and it usually suits all my models (as a concept): simpler model > better fitting my training but more risk of overfitting.

In your case, if you find that a more complex model performs worse unless it is because of specific reasons connected with the model structure (like it misses some pieces. In this case, it doesn't, compared to a compartmental model). You have broad enough priors; I would say this depends on the characteristics of the training vs. validation datasets. I imagine your more complex model is overfitting; in other terms, it ends up accommodating the information content of the training too much and loses generalization power since, for some reason, your validation has different information, like some regular bias (which, for soil, is not particularly surprising).

AC4: 'Reply on RC2', Konstantin Aiteew, 18 Aug 2023

The difference between MONICA and RothC (and other simple SOC models used here) is that MONICA quantifies the carbon input from crops by crop growth modelling (which is also affected by simulated nitrogen cycling, soil water balance and other processes) while for RothC we used observed yields to estimate carbon inputs from crop residues.

Because measured carbon inputs from crop residues were not available, we used observed yields to get an indication of the ability of MONICA to simulate crop development and finally carbon inputs. Although we were able to adequately replicate yields (just as a proxy of modelling carbon inputs from crop residues) some deviances between model and observed yields remain, so that the description of variability between sites is not perfect. One reason here is that some processes are not described by the model (the effect of phosphorus and other nutrients or the effect of crop diseases).

Contrary, the carbon input estimation method used to derive carbon inputs for RothC. Here we used directly observed crop yields and allometric functions, evaluated for Germany in previous studies (Dechow et al. 2019, Riggers et al 2019).

Like you mentioned, there is not much difference between decomposition approaches of RothC and MONICA. Therefore, no big differences between models would have been expected. However, we assume that a less good description of carbon input variability between sites by MONICA cannot be compensated by more degrees of freedom during the calibration process, because the carbon input estimation of RothC is based on measured yields, that should correlate with carbon inputs (Bolinder et al. 2007) and therefore better describe the variability of carbon inputs between sites. We assume, that next steps in improving the description of soil organic carbon development, should address crop growth modelling and carbon input estimation by dynamic models. Considering the several dozen parameters already associated with crops in MONICA, this is quite challenging. So, in the end the performance is limited by the working hours that can be invested in the model and the data availability. A simple model has the advantage that it takes significantly less time to fully exploit the potential of the model.

Based on that, it is probable that with more data and time invested in the crop growth model in MONICA a better performance could be achieved in the SOM decomposition submodel.

Dechow, R., Franko, U., Kätterer, T., and Kolbe, H.: Evaluation of the RothC model as a prognostic tool for the prediction of SOC trends in response to management practices on arable land, Geoderma, 337, 463-478, https://doi.org/10.1016/j.geoderma.2018.10.001, 2019.

Riggers, C., Poeplau, C., Don, A., Frühauf, C., and Dechow, R.: How much carbon input is required to preserve or increase projected soil organic carbon stocks in German croplands under climate change?, Plant and Soil, 460, 417-433, 10.1007/s11104-020-04806-8, 2021.

RC2: 'Reply on AC2', Lorenzo Menichetti, 07 Aug 2023

If with an UNCONSTRAINED hill climbing algorithm, you get worse results from a more complex model; I would say that unless you just hit some local optima and the algorithm got stuck in it (for testing this, try starting from different initial values), then it should be either because of a bias between training/validation or some structural reasons.

Was it an unconstrained calibration, in that case? Or did you still specify limits for the parameter ranges?

AC4: 'Reply on RC2', Konstantin Aiteew, 18 Aug 2023

Regarding the hill climbing algorithm. We used a constrained hill climbing algorithm, were we defined specific parameter ranges. We started the algorithm several times with different initial values. It came handy as a relatively simple calibration method that allowed us to compare the performance of the Bayesian calibration.

RC2: 'Reply on AC2', Lorenzo Menichetti, 07 Aug 2023

A compartmental SOC decomposition model has a few possible structural reasons, particularly for models with more than 3-4 pools (usually enough to describe most SOC dynamics). One is model initialization: if, for some reason, one of your models is getting too much slower pools, this will, of course, impact predictions. In this case, though, you could find heteroskedasticity in the errors over time since this model bias would differ. Another one is different forcing functions: the main difference between first-order compartmental SOC decomposition models is in the reduction functions, the rescaling of the kinetics with temperature or moisture or some other edaphic property. For the rest, they are all very similar. It could be an interesting discussion to go deeper into this topic, checking (provided that the models are initialized in the same way) possible differences in the model response functions. Maybe one of your models has better climate or edaphic scaling than others, which would explain fitness differences.

AC4: 'Reply on RC2', Konstantin Aiteew, 18 Aug 2023

We agree, beside carbon input estimation methods, model initialization and response function might be another reason for different model behavior. In the case of soil decomposition, models come from different regions in the world representing different training data sets, which will affect their sensitivities in other regions of the world. Our data set of long-term observational sites is probably not suitable to analyze the effects of model initialsiation or environmental conditions in more detail. This needs data sets with well-defined conditions (indications of initial pool distributions, carbon inputs, root exudates). Here it would be interesting to force carbon inputs in MONICA by observed (estimated) carbon inputs to make decomposition approaches comparable.

Currently, MONICA uses a standard distribution for the pools, which can be considered as a weak point of the SOC submodel. Compared with RothC, which can be run in equilibrium mode or short-term mode (with pool distribution based on radiocarbon values). It would be great if it was possible to adjust the pools in MONICA, e.g. based on radiocarbon measurements or land-use history.

RC3: 'Comment on egusphere-2023-760', Livia Rasche, 10 Oct 2023
Agriculture on organic soils is major source of emissions, but crop models often perform better on mineral soils or are not equipped to simulate dynamics of organic soils at all. How does MONICA handle the representation of organic soils?

AC6: 'Reply on RC3', Konstantin Aiteew, 12 Oct 2023

We agree that most current process-based models were developed specifically to represent mineral soils. The MONICA model is no exception. The nutrient dynamics, mineralisation, crop types, water balances and other factors are specifically adapted and calibrated for mineral soils. However, with the advent of reliable hydraulic parameters for organic soils as published by Wallor et al. (2018) in Geoderma (Geoderma 313, 69-81 and 319, 208-218) MONICA would in principle be able to simulate water dynamics in organic soils. The carbon and nitrogen turnover routines in MONICA are also in principle capable of reproducing the C and N dynamics in organic soils; the influence of the water table on the decomposition processes is currently under revision (Khaledi et al., J. Veg. Sci, under review). At this stage, MONICA has not been tested for simulating organic soils, but also this is currently under development (ongoing PhD project).

RC3: 'Comment on egusphere-2023-760', Livia Rasche, 10 Oct 2023

Why were the years 1992 to 2000 removed from the analysis?

AC6: 'Reply on RC3', Konstantin Aiteew, 12 Oct 2023

We removed the years 1992 to 2000 from the analysis to avoid statistical biases. The initial soil organic carbon measurements were used to initialise the model (initial SOC content is a necessary model input). At the beginning of the simulation, the measured soil carbon content is not different from the modeled one. If we had left all years in the statistical analysis, this would have resulted in a better model performance due to overfitting. Furthermore, in most cases, soil organic carbon changes very slowly. As a result, the actual model performance can only be adequately compared with a correspondingly long series of measurements, and from our perspective, removing the first years of measurements provides a more accurate representation since the measured and simulated values do not coincide as likely.

RC3: 'Comment on egusphere-2023-760', Livia Rasche, 10 Oct 2023
Should an explicit microbial model (Chandel, Jiag,Luo (2023), https://doi.org/10.1029/2023JG007436) be included for a better representation, or do you think that the current approach is sufficient?

AC6: 'Reply on RC3', Konstantin Aiteew, 12 Oct 2023

Using independent data for validation, we have shown that the calibrated MONICA model can describe soil carbon changes within a time frame of 25 years under temperate climate conditions. SOC models that explicitly model microbial dynamics might be advantageous for representing processes like priming or SOC responses on warming. However, the microbial models reviewed in Chandel et al. 2023 vary widely in terms of model structure, processes considered and parameters required, which makes it all the clearer that there is currently no consensus on which approach is best suited to represent SOC

decomposition. Some of the models are evaluated and compared in Sulman et al. (2018) in Biogeochemistry (141, 109-123) using experimental data from litter input studies and warming experiments. In evaluating these models, they find high variability in model results and conclude that first-order models already produce divergent projections due to parameter uncertainties, and that structural diversity among models would exacerbate these uncertainty. We believe that increasing differences between models due to structural diversity does not represent a degradation in predictive ability, but rather a more accurate estimate of predictive uncertainty (Bradford et al., 2016 in J. Ecol 104, 229-238; Lovenduski and Bonan, 2017 in Environ. Res. Lett. 12). It is unclear if the adoption of one of the various microbial approaches could increase the model accuracy of MONICA. Because of the increased model complexity and number of parameters, the evaluation and calibration of an incorporated microbial model for regional conditions, would also require a more constraining data set, preferably including those events where microbial models might outperform models based on first order kinetics.

RC3: 'Comment on egusphere-2023-760', Livia Rasche, 10 Oct 2023
Do you think that the model would perform robustly on sites without sufficient field data for calibration, or on larger scales?

AC6: 'Reply on RC3', Konstantin Aiteew, 12 Oct 2023

As mentioned earlier, we evaluated the calibrated model using independent data (sites). Therefore. we believe MONICA provides robust estimates of SOC changes for temperate climates, a wide range of soil conditions, and management options represented in the data set (crops, crop sequences, organic amendments etc.). How MONICA performs with data outside this definitional range is unknown to us and should be investigated through further model testing. However, the model's ability to describe plant growth, which is necessary for estimating carbon inputs from crop residues, has been extensively tested for a wide range of climate and soil conditions (Bassu et al., 2014 in Glob. Change Biol. 20, 2301-2320; Kostková et al., 2021 in Agri. Sci. 159, 69-89, …). Currently there are projects underway that investigate the validity of MONICA simulations for crop yields and soil carbon changes at different spatial scales, but for now we are confident that MONICA can provide adequate results for a variety of different cropping systems in temperate climates.

Changes made in the manuscript.

1. We changed the affiliations of Prof. Dr. Claas Nendel.
2. At line 51 we go more into detail about the difficulties of complex models.
3. At line 108 we added one additional references, which relate to the soil temperature and moisture data we used in the study.
4. At line 230 we added one sentence, which explains in more detail how we have chosen the prior probability distribution of the parameters.
5. At line 236 we go more into detail, how subsequent parameter value changes were selected during the parameter calibration.
6. At line 281 we go more into detail, why we removed the years 1992 to 2000 from the statistical analysis.
7. At line 453 we go more into detail why it is currently difficult to include a more explicit microbial model in the MONICA model.

8. We changed the data availability section, describing in more detail how to obtain the data from the state government of Lower Saxony and the German weather forecast service (Deutsche Wetterdienst) and updating the link to the repository to obtain the openly available data.
9. We changed the code availability section, updating the link to the repository.
10. We removed the software availability section.
11. We included new acknowledgments in the acknowledgments section.