Author’s response

Dear Boris Jansen,

Thank you for handling our submission to SOIL. We have finalized the revised manuscript based on the referee’s comments. We incorporated most comments from the referees, with the exceptions of the ones where we indicated this in the rebuttals. A copy of the rebuttals is attached below. With the following points we summarize the most important modifications to the manuscript:

1. We added a new Appendix that described the SOC cycle of the SoilGen model in detail. It provides flowcharts, equations and parametrization that supports the model results provided in the paper.
2. We better motivated our choice of protection mechanisms and the choice for the implementations and range of protection values (lines 128-136 & Appendix A). We re-emphasized that these implementations were based on first academic insights and require more empirical evidence to refine their implementation and calibration.
3. We better supported the model parametrization and calibration using the relevant literature (lines 97-99, Appendix A) and did a reality check on the magnitudes of the model output (lines 385-394)
4. We explained the regression analysis in more detail, with a more extensive description of what the different parameters represent. We summarized this in the new regression equation (Eq. 2). We excluded simulation years 0-1000 from the regression to avoid effects of model-warm up in the analysis.

With these modifications, we feel that the manuscript has improved in clarity and will have a greater impact. We’re curious to hear what you and the reviewers think of the changes.

We have one remaining question regarding the Appendices. We have included the description of the SOC module (Appendix A) and additional figures (Appendix B) as appendices to the paper, as we believe that this material is best presented there for easier reference. However, they have become quite substantial, increasing the overall length of the paper. We therefore wanted to ask whether the journal prefers this type of material as an Appendix, or whether it would be more appropriate to include it as a separate Supplement.

With best regards,

Marijn van der Meij

Rebuttal referee 1

This paper presents the results of scenario simulations with the latest version of the soil evolution model SoilGen, which considers four different mechanisms that influence the strength of protection of soil organic carbon (SOC) against microbial decomposition. Many properties that are considered as static and therefore described with constant parameter values in almost all other models of transport processes in soil, are treated as dynamic variables in the SoilGen model. This paper therefore fits very well to the topic of the special issue in SOIL ("Advances in dynamic soil modelling across scales")

Response: Thank you for the kind words and constructive review. Your comments will be helpful to improve the clarity and completeness of our manuscript. Note that we numbered the comments for easier reference.

1. *The simulations can be seen as a kind of sensitivity analysis to investigate the relative importance of these different stabilization mechanisms, along with environmental factors like climate and land use, for soil organic carbon sequestration in a very long-term perspective (millenia). Although strict validation of such model predictions has not been attempted (it seems very difficult for such temporal scales), the authors have tried to do some "reality checks" in section 4.1.3, for example with respect to the effects of erosion and afforestation on C-stocks. Could this be further developed? I was thinking, for example, of the estimates of C-stock loss due to the development of intensive agriculture over longer time spans than centuries, given by Lal (2013) and Sanderman et al (2017).*

Response: We will extend our reality checks of the simulation results. In fact, we already simulated the effects of the continuation of intensive agriculture. The results are provided in Figure 5 as the continuation of agriculture without rock additions, but we have to admit these results are not clearly presented. We will modify Figure 5 to better visualize the effects of ongoing intensive agriculture by 1) extending the X axis from the end of the natural phase (year 8000) until the end of the projected SOC changes (year 12000), and add the line for the natural reference in each subplot.

These reference levels indicate a loss of 45% relative to the antecedent SOC pool by intensive agriculture, which fits with the ranges reported by Lal (2013; 25-75%) and Sanderman et al. (2017; 14-61%). The average values of the latter are somewhat lower than ours (14-26%), but that can be due to their very small sample and global representation. In Sanderman et al. (2017), it is also not clear if their results describe SOC-loss for a new steady state situation after land use conversion, which would implicate that their values for SOC-loss may be too low. We will add this analysis to our reality checks in Section 4.1.

2. *The description of the model itself is very brief. Some more detail would help the reader, particularly on the organic carbon turnover model.*
    a. *For example, in equation 1, could you explain how the "crop cover" modifier is estimated? What aspects of crop cover influence SOC decomposition? How does it relate to the land use scenarios shown in table 2?*
    b. *Apart from re-distributing SOC in the plough layer, does tillage also affect decomposition rates?*

Response: We deliberately refer to the book on SoilGen3.8 for a detailed description, to prevent a very extensive paper. However, we agree that a more extensive description of the SOC module and its modifiers would be helpful for the readers. That is why we will add an Appendix or supplement describing these model processes to the paper.
Below, we reply briefly to your specific comments. More extensive descriptions will be added to the Appendix. The SOC module and its modifiers are taken from the RothC26.3 model, but their drivers are mechanisms simulated by SoilGen itself. The different modifiers affect the decay rate constants of the SOC module.

Crop cover is added as a modifier, because SOC decomposition has been found to be faster in fallow soil compared to cropped soil. For agricultural crops, the crop cover increases from 0 at germination date to 100% at harvest time via an exponential function. Natural vegetations are assumed to have a constant crop cover of 100%.

Temperature is calculated every 0.1 day for each soil compartment by a heat flow routine driven by atmospheric temperature and a volumetric heat capacity responding to the volumetric water content. A higher temperature will speed up decomposition.
The **moisture deficit** responds to the water flow routines (timesteps variable but smaller than 0.1 day) and the precipitation deficit, and is corrected for the clay content. A large moisture deficit will slow down decomposition.
Tillage does affect decomposition rates as follows: by mixing, the texture and SOC will change in each compartment, affecting the h-θ-K relations and the bulk density. This will influence heat flow and water flow, impacting the temperature and moisture deficit modifiers. Additionally, the geochemical modifiers may respond to mixing as well, for example via calcite content, ground rock content and texture affecting pore size distribution. The depth and intensity of tillage mixing determine which compartments are affected and the rate of their changes. Both depth and intensity of mixing are higher in the intensive-agriculture period. Feedbacks between tillage ad aggregate breakdown are currently not included in the model.

3. *The authors discuss at length in section 4.2. the potential of SoilGen to evolve into a "digital twin", that is to say a model that is data-aware, one that can be updated on receiving new data from the real world. I have to admit that I am a little sceptical about this. I can't see how a model that makes predictions of slow processes for thousands of years into the future could really be a suitable prospect for digital twinning. The gap in time-scales between data collection and model projections seems just too large. So far, dynamic twinning between data and models has only been attempted in short-term studies, as the authors acknowledge at lines 391-395. It would be good if the authors could briefly discuss the disparity in time-scales in the text that follows at lines 396-422.*

Response: It is not so much the potential of SoilGen itself to evolve into a digital twin, but its potential to identify what processes should be in a Topical digital twin focusing on soil organic carbon. Digital twins serve two main purposes: 1) real-time coupling of monitoring data with simulations, and 2) tools for projecting future changes.

Especially for point 2 we need models covering centennial timescales, as these are the scales over which the SOC dynamics will emerge and over which we are actually developing legislation, for example for carbon farming.
Observations over these timescales are scarce.

We will make this distinction between these two purposes clearer in the text and address the disparity between time scales of monitoring and SOC modelling. We will mention two data sources that can be used to benchmark these models: long-term SOC agricultural field experiments and historical records (e.g. López i Losada et al., 2025), and chronosequences of soil development (e.g. Sauer et al., 2009).

4. *On the whole, the paper is well written and nicely presented. However, I think it would be better for the reader if section 2.3 came before the parameterization is presented in 2.2. The scenarios are currently mentioned in 2.2 before they are defined and explained.*

Response: We prefer the original order of these Sections, as 2.2 described shared inputs and parameters between most model runs, while 2.3 explains the differences between the scenarios. The land uses are an exception, but are still shared among different groups of scenarios. We will clarify this in the text.

Specific comments

5. *Line 47: there seems to be a word missing after "environmental"*

Response: Thanks. We will add the word "factors".

6. *Line 49: replace "seem to" by "may". I am not convinced there should be any difference in the mechanisms (only in how strongly they are expressed).*

Good suggestion, we will make the replacement.

7. *Line 51: for the same reasons as at line 49, replace "plays" with "may play"*

Response: Good suggestion, will be replaced.

8. *Line 60: Please update this reference to the final published article: Coucheney, E., Herrmann, A., Jarvis, N. 2025. A simple model of the turnover of organic carbon in a soil profile: model test, parameter identification and sensitivity. SOIL 11, 715-733.*

Response: Thanks, we'll update the reference.

9. *Line 75: perhaps the heading should be: "Model overview"*

Response: Given the small size of this section, "overview" is indeed more appropriate.

10. *Line 79: replace "that include" by "of"*

Response: Good suggestion, will be replaced.

11. *Line 82: "seldom" not "seldomly"*

Response: You're right, will be replaced.

12. *Line 88: "have undergone" not "underwent"*

Response: Sounds better indeed, will be replaced.

13. *Line 105: Instead of "default rate", it would be better to write "reference rate constant". Strictly, R is a rate constant, not a rate. It would be good to make this clear by writing the units of R here (as well as the other terms in equation 1).*

Response: Agreed. We will check and replace throughout the text. R is in y-1.

14. *Line 132: Comparison of model output with what? This should be clarified.*

Response: "model output" will be changed to "model outputs for the various scenarios".

15. *Lines 177-179: Is this really a model artefact?*

Response: This is indeed not an artefact, but an effect of how the erosion process is implemented. We will change "model artefact" to "implementation effect".

16. *Lines 353: This comparison seems a little dubious to me because no-till can affect SOC stocks through several other processes, not just by reducing rates of erosion. Surface crop residues can modify the soil thermal regime, which in the decadal time perspective that is relevant here, could be just as important as erosion. Tillage also enhances decomposition by disrupting soil aggregates. It's not clear to me whether SoilGen considers these processes or not. This should be clarified.*

Response: Indeed, no erosion is not the same as no-till. All our agricultural scenarios do include tillage, with shallow plowing for extensive or deeper plowing for intensive agriculture. Next to that, the agricultural scenarios also differ in the net OC input.
The effects of tillage on decomposition in SoilGen will be described in the new Appendix on the SOC module. The disruption of aggregates itself is, probably unjustified, currently not a factor in SoilGen.
We will add a remark on the limitations of SoilGen with regard to tillage affecting decomposition.

17. *Lines 365-366: "ground" or "crushed" not "grinded"*

Response: We'll use "ground".

References

Lal, R.: Intensive Agriculture and the Soil Carbon Pool, Journal of Crop Improvement, 27, 735–751, https://doi.org/10.1080/15427528.2013.845053, 2013.

López i Losada, R., Hedlund, K., Haddaway, N. R., Sahlin, U., Jackson, L. E., Kätterer, T., Lugato, E., Jørgensen, H. B., and Isberg, P.-E.: Synergistic effects of multiple “good agricultural practices” for promoting organic carbon in soils: A systematic review of long-term experiments, Ambio, 54, 1715–1728, https://doi.org/10.1007/s13280-025-02188-8, 2025.

Sanderman, J., Hengl, T., and Fiske, G. J.: Soil carbon debt of 12,000 years of human land use, PNAS, 114, 9575–9580, https://doi.org/10.1073/pnas.1706103114, 2017.

Sauer, D., Schülli-Maurer, I., Sperstad, R., Sørensen, R., and Stahr, K.: Albeluvisol development with time in loamy marine sediments of southern Norway, Quaternary International, 209, 31–43, https://doi.org/10.1016/j.quaint.2008.09.007, 2009.

Rebuttal referee 2

In their manuscript, titled 'Modelling long-term soil organic carbon sequestration under varying environmental drivers and internal protection mechanisms – towards a digital twin', van der Meij and Finke describe their result of a study on the relation between long-term soil organic carbon (SOC) dynamics and soil development, using the SoilGen model.

The main aim was to assess the effect of three environmental drivers (bioclimate, soil erosion and land use) and four SOC protection mechanisms (soil aggregation, clay mineralogy, microporosity and metal oxyhydroxides) on long-term SOC dynamics. To do so, the authors ran SoilGen for artificial European loess soils in two different climatic settings (a dry natural steppe vegetation and a wet natural deciduous forest) for a time period of 10,000 years, with an additional 2,000 years to simulate recovery.

Their model results showed that the most important protection mechanism of SOC varied temporally depending on how soil development proceeded, with stabilization of SOC in aggregates initially being the most important protection mechanism. Their simulations showed that bioclimate had the largest influence on SOC, while protection mechanisms the lowest. In addition, the model showed that the subsoil recovered slower than the topsoil after an intense agricultural phase.

It is well-know that there is an important link between climate, soil development, mineralogy and soil organic matter (SOM) dynamics. Therefore, linking these processes in numerical models is an important step towards improving the simulation of SOM at the landscape scale, or over long time scales. This study is therefore timely, and tackles an important topic that has largely been overlooked in modelling studies.

The manuscript is well written and the results are clearly presented. I think it is an important contribution to the field over SOC – soil development modelling. I hope my feedback can improve the quality of the manuscript, and make some aspects more clear to the reader.

Thanks for your extensive comments and valuable feedback on our manuscript. Below, we provide our responses to your comments. Please note that we have numbered your comments for easier reference.

My main feedback concerning the present manuscript is the following:

1. The statements the authors make about feedbacks between soil development and soil organic carbon can be more nuanced or more extensively compared to literature, as these are based on model results that have not been validated.

Response: We think you are referring here to Section 4.1.1. We will extend this Section with additional references to support the evaluation of our model results.

The model SoilGen is actually validated in numerous other studies, focusing on specific processes (e.g. Finke et al., 2015) or feedbacks between processes (e.g. Ranathunga et al., 2021). These studies give us the confidence that the SoilGen results are at least plausible. We will refer to additional SoilGen papers to show its validation status.

2. Reporting of the turnover times of different simulated SOC pools would be valuable information for the reader to interpret the results.

Response: We will add a new Appendix that will describe the SOC module in more detail, including the reference turnover rates of the different SOC pools. For the sake of completeness, we report them here as well: kRPM=0.35 a-1; kDPM=10 a-1; hHUM=0.008 a-1; kBIO=0.66 a-1

3. The description of how the RothC pools are linked to the protection mechanisms should be more detailed. Which RothC pools could be protected by aggregates, MOOHs, clay minerals and pores? This information is necessary for the reader to understand how the model functions.

Response: The new Appendix describes the links between the different SOC pools and the effects of the rate modifiers.

All four SOC pools have a unique reference rate constant, obtained by averaging point-calibration results in other studies in loess soils (Yu et al., 2013; Ranathunga et al., 2022). All four pools are similarly affected by the different rate modifiers as described in Eq. (1) in the paper. Humus pools (84%) and Resistant Plant Material pools (14%) dominate the SOC because their lifetime is longest, and therefore these pools are protected the most from the protection mechanisms in absolute terms.

We'll add a statement on the equal treatment of all pools, and the relative importance of these pools.

4. Some information about the methods is missing, for example:

a. Was there a different parameterization for processes in the topsoil and subsoil?

Response: There was no different parametrization. We will add a sentence on this issue to avoid misinterpretations.

b. Did SOC turn over slower in the subsoil? If so, how was this simulated?

Response: We did not look into specific turnover times for top- and subsoils, but the effects of the modifiers on the reference rate constants (eq.1) will differ along the soil profile due to the vertical distribution of moisture, clay, calcite/base saturation, MOOHs and resulting porosity. This vertical distribution varies, among others, by climate, erosion level, plowing depth and time of year. We will also clarify this point to avoid misinterpretations.

c. Which depth layers were simulated?

Response: We divided the soil profiles into compartments of 5 cm. The initial soil thicknesses differed between the erosion scenarios and were chosen to preserve a soil thickness of at least 150 at the end of the simulations. We will add this information to Section 2.2.

d. See more detailed aspects below.

5. More information on the statistical analyses is needed for the reader to understand how these were performed.

a. Why was a linear model used? Were all relations linear?

Response: Most predictors were categorical, and therefore no linear relationship between these predictors and SOC stocks is required. In a linear model, categorical predictors simply estimate differences in mean SOC stocks between groups. Only the simulation year was treated as a numerical predictor. Although simulation year did not show a linear relationship with SOC stocks on its own, we still included it in the model to account for any linear trend in combination with the categorical predictors.

Linear regression is the appropriate method here, because it estimates differences between groups and provides the basis for the ANOVA, which we used to assess how much each predictor contributed to the overall variation in SOC stocks.

Based on your comments, we re-evaluated the diagnostic plots and saw that the assumptions of normally distributed residuals and homoscedasticity were violated. This was caused by data from the start-up phase of the model, where the SOC stocks were still very low and needed to reach the natural equilibrium. In the revised manuscript, we will filter out all simulations years below 1000 to remove this effect. Re-evaluation of the residuals shows a distribution that is much closer to normal and a better fit of the linear model (R2 = 0.96), improving our confidence in parameter standard errors and results of the ANOVA. As a consequence, the effect of simulation year became 100 times smaller than in the initial linear regression, indicating that the categorical predictors dominated the SOC stocks.

We will clarify the use of categorical predictors in the linear regression models in the manuscript.

b. What was regressed against what?

Response: This information was already provided in lines 185-187 of the methods, but we will clarify this by writing out the regression equation in the revised manuscript.

c. Which data was exactly used in the statistics? Data for every simulated year?

Response: This was also already mentioned in the manuscript (lines 185-187): "Bioclimate and protection mechanism were treated as categorical variables that <u>varied between scenarios</u>, while erosion class and land use were categorical variables <u>specified at the yearly level</u>, reflecting changes in the erosion intensity and land use throughout the simulations. The simulation year itself was included as a continuous numerical variable". We will clarify this in the regression equation by indicating which variables are time-dependent and which were constant for each scenario.

6. The figures can be made more clear:

a. Avoid using abbreviations in the titles, or explain them in the caption. For example, mention the different erosion intensities, instead of E0, E1, E1E2E3, etc.

Response: We will clarify the abbreviations of the erosion scenarios in the Figures:

- Figure 1 and 3: we will mention the actual erosion rates in the titles and on the axes.
- Figure 2 and 4: we will mention the rates in the caption and refer to Table 2 where they are written out.

b. A box around each plot would make it easier to distinguish between plots.

Response: We prefer our plots without additional boxes or separations, as the additional lines will distract from the contents of the plot. We will increase the white spaces between the plots for visual separation.

In my feedback, I mention certain published articles. These have been chosen based on their scientific relevance, and I leave it up to the authors whether they want to include these in their manuscript or not.

Specific comments

--- Introduction ---

7. Title: Why is the term 'internal' protection mechanisms used, also throughout the manuscript? As this term isn't commonly used in the SOM literature, perhaps just talk about 'protection mechanisms'?

Response: We used "internal" and "external" to distinguish between internal soil processes and feedbacks affecting degradation of SOC (protection mechanisms) and external environmental conditions at the soil boundary (climate, vegetation and land use). We prefer to keep this distinction to facilitate the discussion of the drivers of the SOC cycle, but we will clarify this in the Introduction.

8. L23-24: as the terms 'full digital twins' and 'topical digital twins' have not been introduced, the reader will not know what you mean by this.

Response: We will clarify the difference between a full and topical digital twin in the abstract, following the definitions we provide in the manuscript (lines 38-39, 389-393).

9. L31-32: The study by Minasny et al. (2017) led to a large scientific debate in the literature, as many scientists did not agree with their outcomes. I would encourage the authors to also mention some of these studies, to show both sides of this debate (for example: Schlesinger and Amundson, 2019 (https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14478); van Groenigen et al., 2017 (https://doi.org/10.1021/acs.est.7b01427); de Vries, 2018 (https://doi.org/10.1016/j.geoderma.2017.05.023); VandenBygaart, 2018 (https://doi.org/10.1016/j.geoderma.2017.05.024); Poulton et al., 2018 (https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14066))

Response: We refer to Minasny et al. (2017) to support our statement "Enhanced carbon sequestration in soils could partly or completely counterbalance global anthropogenic carbon emissions". There is no debate about this rather wide statement and we think this reference is suitable, because it increased the awareness of soil carbon sequestration potential.

In our manuscript, we do not take a position in the feasibility of the 4-per-mille ambition, as it sidetracks from what we have been investigating. Therefore, we will not include the suggested references that will trigger this debate.

10. L39: The term 'digital twin' can be described more extensively here, so the reader knows what you mean.

Response: We will extend our description of digital twins in the introduction, while avoiding duplicating text from Section 4.2.

11. L55: it would be useful to have a better introduction on soil evolution models. Are there alternatives to SoilGen? How about trade-offs between process representation in a complex model like SoilGen, and data availability? How well constrained are the parameters of SoilGen based on previous studies

Response: We intendedly focused our introduction of soil evolution models on those simulating soil organic carbon dynamics, as that is the topic of the paper. We will add a few additional sentences on soil evolution models in general and provide two references that give an overview of different models and their process coverage (Van der Meij et al., 2018; Meng et al., 2022).

To our knowledge, there are no alternatives to SoilGen for modelling soil genesis on a profile scale. Most soil evolution models simulating SOC dynamics are single-purpose, while SoilGen simulates many additional soil processes and uses them as drivers for the SOC cycle. We will clarify this in the text and will refer to a recent book on SoilGen3.8 (Finke, 2024), that describes different parametrization, calibration, verification and application studies. We feel that we would overload the manuscript by adding all those studies to the introduction. Studies specifically focusing on the SOC cycle will be cited in the new Appendix describing the SOC cycle in SoilGen.

12. L67: define what you mean by 'long timescales'.

Response: We will our definition of these timescales (centuries – millennia).

--- Methods ---

13. L83: Were cover crops simulated?

Response: No, because during the majority of simulation time cover crops were not applied. The standard crop is Barley, emerging in March, mature (crop cover and roots) in May and harvested in late summer. We will specify crop type in Section 2.2.

14. L87-91: it would be useful for the reader to mention in which climates and for which soils these previous SoilGen analyses were performed. That way, the reader knows if that's relevant for this study or not.

Response: SoilGen has been tested and applied in a variety of settings and climates, including the ones simulated in this study (loess, temperate and semi-arid climates, natural and agricultural land use). We will mention this in the manuscript and refer to the SoilGen book for an extensive overview (Finke, 2024).

15. Section 2.1.3: it would be good to also explain what the 'no protection scenario' is, and how the reader should interpret this.

Response: We will clarify that "No protection" means that there is no modified decomposition rate due to one of the listed protection mechanisms. In other words, the rate modifier has a value of 1.

16. Eq 1: it would be better to provide intuitive letters, instead of abcd

Response: The letters a, b and c are used as the rate modifying factors for temperature, moisture and crop cover in the RothC26.4 model (Coleman and Jenkinson, 1996). As the SOC module is based on RothC, we think it is appropriate to use the same terminology, and, in the same line, add the additional rate modifier for protection by hydrogeochemistry with the letter d.

17. Caption Table 1: Why are the rate modifiers limited to values between 0.8 and 1? Please explain this in the text

Response: Although there are many studies showing that there are different hydrogeochemical protection mechanisms for SOC decomposition, there are no robust estimates of the magnitude of this effect. We decided on a range of 0.8 to 1.0 because of the lack of published research to get better values. This is also stated in line 125: "It should be noted that these modifiers are currently loosely based on first academic insights in these processes".

We also decided to give each protection mechanism the same range of values to give them equal potential effects. We will add a statement supporting our choice for equal values for each mechanism and address the uncertainty of these values and the need for studies constraining them in the discussion

18. Table 1: it would be much more intuitive for the reader if graphs are provided that show how these equations affect the rate modifiers. This is very difficult based on the equations alone, and as this is central to the performed simulations, the reader would benefit from a better understanding of this.

Response: We will provide a graph visualizing the effect of the rate modifiers in the new Appendix describing the SOC cycle.

19. L137-140: Can the values for NPP that were used in the simulations be provided?

Response: We will add values for NPPs for the different land uses for both climates to Table 2. Note that SoilGen corrects NPP for yield loss and growth while calculating the net C-input by litter to the soil, following the methodology provided by Sallaba et al. (2015) as explained in the manuscript, which will be further explained in the new SOC appendix.

20. L143: It would be good to explain how the C inputs were vertically distributed during the simulations.

Response: The endorganic litter (decaying root material) is distributed with depth following the root density distribution, while the ectorganic litter (leaf material) is added on top of the first soil compartment. Both types of litter are subsequently redistributed by plowing and bioturbation. We will specify this in the text.

21. L145-146: Please explain how soil material was mixed through bioturbation. Was this magnitude constant with depth, or exponentially decreasing? What was the ploughing depth?

Response: Bioturbation was not constant over depth. Over the first 25 cm, it is (annually) 0.3% of the mass, from 25 to 40 cm it decreases linearly with depth from 0.3% to 0%. Together this results in mixing of ca. 11.9 Mg ha-1 a-1 (l.145).

Ploughing depth was 10 cm in the extensive agriculture period, and 25 cm in the intensive agricultural phase. In the plough layer, 50% of the mass is mixed during each plow event.

We will clarify this in the text and update Table 2, where the plowing depths were reported wrong.

22. Table 2: was $CaCO_3$ added to the soil surface, or vertically distributed along the soil profile?

Response: $CaCO_3$ was added to the surface in the autumn, but rapidly redistributed over the plow layer due to annual plowing outside of the growing season. There is also some shallow leaching during winter, but effectively there is equal vertical distribution over the plowed layer when the growing season starts. We will clarify this in the manuscript.

23. L154: What were the values of the average July and January temperatures used in the simulations?

Response: In the dry bioclimate, average winter and summer temperatures were 2 and 20 °C, while the dry bioclimate had average winter and summer temperatures of -4 and 20 °C. Both climates had constant average temperatures over time. We will add this information to Table 2.

24. L173: Please mention the thickness of this topsoil layer

Response: The thickness of each soil compartment, including the top layer, is 5 cm. We will mention this in Section 2.2.

--- Results ---

25. L198: I would rather talk about 'higher compared to the base scenario', instead of 'increase', as the latter implies an increase over time when talking about SOC dynamics.

Response: We will modify the text following your suggestion.

26. L220-221: it would be more useful to mention if NPP was the same between these sites, in addition to that they have the same vegetation

Response: The NPPs are almost the same, but there can be a max 2% deviation due to the effects of temperature and precipitation in the calculation of NPP. Following our reply to comment 19, we will report NPPs in Table 2.

27. L221: It would be useful to explain why the decrease in SOC after 8000 years is larger in the dry compared to the wet scenarios.

Response: The decrease is larger because the initial natural SOC stock is larger. We will mention this in the text.

28. L223: it would be useful to already mention this frequency of layer removal in the methods, when erosion in described.

Response: The frequency of layer removal follows from the erosion rates and layer thickness, as mentioned in Section 2.4. We will specify this more clearly in the text and explicitly mention the frequency for the different erosion scenarios.

29. Figure 3: It is not clear based on which data points the box plots were constructed. Were these the C values for every year in the same simulation between 9,900 – 10,000?

Response: Correct, each box in the plot shows the distribution of SOC values from the simulation years 9900 – 10000 for one specific scenario, which is also reported in the Caption of the Figure. We will specify that these boxplots are comprised of annual C stocks.

30. L247: Is this because the simulations were started with very little C, which increased over time? And since there were different forcings at different points in time, would time not by definition have an effect on SOC stocks? This needs more explanation.

Response: Most of the variation in SOC stocks can be explained by the other categorical predictors. These are partly time-dependent, as mentioned in our response to comment 5c. Within the remaining variation, there is only a very small part that can be explained by the simulation year, especially after excluding the first 1000 years of simulation (comment 5a). We will clarify the time-dependence of different predictors and explain why the simulation year itself only plays a minor role.

31. L264-265: It would be good to mention the turnover rates of the pools that were used, as these values are determined by those parameters.

Response: See our response to comment 2.

32. Caption Fig. 4: would be good to mention that the lines connect dots at the same erosion rate

Response: These lines do not connect the same erosion rate, but different erosion rates within one bioclimate (point shapes) and soil compartment (line types), which is also defined in the legend. We will clarify this in the caption.

33. Figure 4: How are topsoil and subsoil defined? This is not mentioned in the manuscript?

Response: The definitions are already provided in two places: lines 181-183 in the Methods and the Caption of Figure 4. Topsoil: 0-25 cm. Subsoil: 25-100 cm.

34. Figure 5: It would be more intuitive if all y-axes had the same range

Response: The varying axis ranges enable the visualization of small differences between the lines, but following comment 1 of the other referee, we will modify this Figure by showing simulations results from years 8000 – 12000 to facilitate comparison with the natural reference. As a consequence, the plots representing one bioclimate will get the same Y axis ranges.

--- Discussion ---

35. L297-299: You state that 'results revealed surprising insights into SOC dynamics' and 'the model can advance our understanding of these feedback and drivers'. However, the model results have not been confronted to data, and it seems you relied on model parameters determined in previous studies. Please support these statements by explaining why you have confidence in the correctness of these results.

Response: This is a recurring comment in (soil) modelling studies such as ours. Models such as SoilGen represent our current state-of-knowledge of how soil processes function, and consequently how soils evolve over long timescales. These models are continuously improving, following new academic insights and calibration and verification studies, as listed in several parts of the manuscript (e.g. lines 85-100). The fact that SoilGen explicitly simulates the soil water and heat cycles as drivers of soil processes, makes it possible to apply the model in a range of soil types and geographical settings (lines 325-326).

We don't claim that the model results accurately represent a specific field site (lines 320, 339-344), but with the reality checks throughout Sections 4.1.2 and 4.1.3 we show that the model provides plausible results. We also stress that calibration is required for deriving accurate estimates of site-specific SOC sequestration (line 342-344).

The validation status of the model and plausibility of the results gives us the confidence that SoilGen accurately simulates soil processes in our simplified scenarios and that emerging feedbacks can actually occur in soil systems. This explorative modelling differs from reproducing specific field settings (Fig. 5 in Temme et al., 2017), but provides its own kind of valuable information about soil system functioning and feedbacks.

We hope that this reply explains our confidence in the model. Following comment 1 of the other referee, we will extend our reality checks of the model results for an even higher confidence and we will refer to Temme et al. (2017) to explain the type of modelling we are performing in this study.

36. L304: Can you say that aggregation is the dominant protection mechanism? Because this protection mechanism is only ca. 10 % larger than the 'no protection' scenario in Fig. 1. Does this mean that ca. 90 % of SOC is in the soil because of the chemical composition of OC, which is the basis on which RothC simulates differences in turnover between the pools? Please explain this in the manuscript.

Response: The protection mechanisms do not just protect the surplus of SOC compared to the baseline, but all SOC in the different pools (see reply to comment 3).

The scenarios with protection by aggregation shows the overall highest SOC stocks, meaning that it reduces SOC decomposition from all the pools the most amongst the protection mechanisms. We also show that the dominant protection mechanism changes over time following decalcification of the soil, as the $CaCO_3$ content is one of the drivers of aggregation in the model (Table 1).

Based on this, and our added explanation of the functioning of the protection mechanisms (comment 3), we think we correctly state that, in a calcareous loess soil, aggregation is temporarily the dominant protection mechanism.

37. L307: 'liming restored aggregation as the dominant protection mechanism': did this take place over the entire soil profile? Or was the CaCO3 only applied at the topsoil and aggregation improved there? It would be useful to the reader to make this clear.

Response: Lime is applied to the surface and redistributed over the topsoil by plowing. As a consequence, the resulting protection mainly affected topsoil SOC. In the dry bioclimate, dissolved $CaCO_3$ reprecipitated at fairly shallow depth, increasing the protection there as well. So, the effect is both in the topsoil and over the entire soil profile under specific climate conditions. We will clarify this in the manuscript.

38. L313-317: This reasoning is not clear to me. What does it mean when 'erosion keeps up with the calcite depletion rate', and why does this ensure that sufficient calcite remains available during the entire simulation? Is there evidence from experiments that minor soil erosion can lead to larger C stocks compared to a non-eroding site? Or is this a model artifact?

Response: This comment comes down to the balance between erosion rates, decalcification rates and SOC sequestration rates. When erosion rates and decalcification rates are similar, the decalcification depth is relatively constant over time. If both rates are slow enough, the SOC cycle can sequester a little bit more carbon due to the added protection provided by the calcite.

We see this effect emerging from other simulation studies as well (e.g. Ranathunga et al., 2021), but we are not aware of experimental studies showing this behavior. We don't consider this a model artefact, since there is a process-based explanation, but we will suggest the need for experimental studies to explore this feedback in the manuscript.

39. L318-319: 'agreeing with observations […]': would be good to have a reference here to back this statement up

Response: We will add (Zhang et al., 2006) as reference to this statement.

40. L332: please explicitly state which soil depths you considered to be subsoil

Response: This is specified in the methods (lines 181-183) and in the Caption of Figure 4, which is referenced *in this sentence, but we will repeat it here for clarity.*

41. L334-335: Please state, here or in the methods, which portion of OC inputs under natural vegetation enters the system as litter, versus as belowground biomass. From these lines it seems as if this was mainly aboveground litter, which is not realistic for forest soils, where C originating from aboveground litter is generally mainly detected in the upper 20 or 30 cm of the soil.

Response: This differs per vegetation type. For the forest, 85% is ectorganic (surface litter) and 15% is endorganic (root litter), following the root distribution. For grasslands, 58% is ectorganic and 42% is endorganic following a more homogeneous root distribution pattern. For agricultural crops, 40% is ectorganic (crop residue) and 60% endorganic.

The ectorganic SOC is vertically redistributed by bioturbation, transporting it to deeper into the soil. We will clarify this in the new Appendix describing the SOC cycle.

42. L335-336: 'agricultural crops contribute more organic carbon directly into the subsoil': do you mean compared to natural forest vegetation? Would you say it's realistic to find larger subsoil SOC stocks under agriculture compared to natural forest vegetation in loess soils?

Response: Based on our model parametrization (see previous comment), there is a relatively larger input of subsurface carbon under agriculture compared to under natural forests. When looking at the succession from cropland to forest, subsoil SOC stocks decrease slightly, while topsoil stocks increase substantially (Figure 5). This is typically a process that needs to be calibrated when the model is applied to field settings, which we will also remark in the manuscript.

43. L340: why 'arbitrary' and 'loosely-based'? Please explain

Response: With these terms, we want to indicate that the simulations are inspired by real geographical settings and earlier parametrizations, but they don't represent a specific study site, see also our response to comment 35.

We do this to fix some of the variation in model inputs, while changing other inputs to assess their effect on model results, similar to a sensitivity analysis. We will clarify this in the text. The term 'arbitrary' is not appropriate and will be rephrased to 'estimated', as the NPP module (Sallaba et al., 2015) estimates the SOC inputs.

44. L340:

a. You mention the turnover rates, but don't mention these in the manuscript. However, this information is needed for the reader to assess if these were realistic, and to know how these changed along the soil profile in the simulations. Please provide this information.

Response: See answer to comment 2.

b. Further, in L342, you say that these 'do not directly represent real-world recovery rates'. How should the reader interpret these? And if this is the case, how can you have confidence in the correctness of the SOC recovery rates you found? Please clarify this in the manuscript.

Response: We state this to mention that we don't simulate a real-world soil, but that we are exploring possible process feedbacks in the SOC cycle of loess soils (see answer to comment 35). The results can be indicative of loess soils, but the actual values will be different for different geographical settings. It is impossible to compare with actual values, as there are no monitoring studies covering these timespans of several centuries.

Based on our confidence in the model, we are also confident that its results are indicative for real-world SOC recovery, but we will emphasize the need for calibration for more accurate site-specific estimates.

45. L347: 'are controlled by external controls' => 'are controlled by external controls in the simulations'

Response: This is not only true in the simulations, but also in soil types occurring in humid climates (Luvisols) versus semi-arid climates (Chernozems). We will add a reference showing the effect of climate on SOC sequestration in these soils and link the bioclimates to specific soil types in Section 2.2: (IUSS Working Group WRB, 2022).

46. L351-354: you state rates of changes in SOC upon land use change or erosion. But higher up, you state that these 'do not directly represent real-world recovery rates'. How does this reconcile? Or am I missing something here?

Response: This is exactly the difference between a sensitivity analysis (what we did) and a site study. Section 4.1.3 shows that the model results are plausible by comparing them to reported values, but the results do not represent a specific soil or study area. We think we were rightfully cautious not to over-interpret or generalize our results and we mainly interpret the trends and patterns.

47. L364-374: A comparison with literature would be valuable here to assess model performance.

Response: We agree that a literature comparison would be valuable (lines 373-374), but we also state that "experimental data on these processes is still scarce". When new experimental data on SOC sequestration by ground rock becomes available, we will gladly compare them with our model outputs to refine process formulations and calibration.

48. L376: Please better discuss why such a digital twin is necessary, and would be preferable to current model approaches. Also, how feasible is this, given the large spatial heterogeneity of the soil system, but spatially and temporarily?

Response: The necessity of soil digital twins arises from the need for quantitative, robust long-term projections of SOC sequestration in soils, for example in the context of carbon farming (European Commission, 2024), as we mention in the Introduction. Digital twins are preferable to current model approaches, as they more accurately represent the current and future soil property states, and can be easily updated when new monitoring data arises.

We think it is more a matter of ambition than feasibility. The mentioned examples worked for specific purposes and sites. We do not make statements on wide soil systems being put in a digital twin, but instead a topical digital twin of a specific soil property. However, there is still a long way to go for developing these robust projection tools.

Based on our response here, we will expand our discussion on the envisaged uses of soil digital twins in the Discussion.

49. L401-402: 'this phase ensures that the model can accurately simulate soil development up to the point where the monitoring starts'. However, above you state that 'the timeline of interest often extends far beyond the period of monitoring'.

Please explain how you can have confidence in simulations that are performed without having the necessary measurements to obtain reliable model parameters.

Response: One of the main purposes of soil evolution models is to develop reliable simulation tools for the long-term development of soils that are widely applicable. There are different datasets we can use for benchmarking these models, not only monitoring data, but also chronosequences (e.g. Sauer et al., 2012), which have been used in different calibration and verification studies of SoilGen.

When we have confidence in the model for simulating soil development from the past to the present, we can also have confidence in its future projections. We feel that we have addressed this sufficiently in other parts of our rebuttal, so we will not repeat these points again in this part of the discussion.

50. L405: Please explain what you mean by monitoring data: which data should be monitored, at which spatial and temporal resolution?

Response: We deliberately kept the description of the topical digital twins general, because the concept can be applied to different processes and properties. The type of monitoring data can be anything related to the process, spatial and temporal resolution and area of interest. Describing this would lead to a large text addition and is outside the scope of this paper.

--- Conclusion ---

51. L424-426: Before stating the results, it would be good to make it clear that these are model results that have not been validated using measurements.

Response: We think we are clear here that we are talking about simulation results, as we mention "synthetic modelling study" in the referenced lines.

Technical comments

Response: We will adopt all technical corrections below in the new version of the manuscript.

52. *L47: something seems to be missing after 'environmental'*

53. *L48: > 25 cm => below 25 cm depth*

54. *L50: Pries et al. 2023 => Hicks Pries et al. 2023*

55. *L52 : 'There is still a lot unknown' => 'A lot it still unknown'*

56. *L87: over => for?*

57. *Caption Table 1: 'protection mechanisms' => 'SOC protection mechanisms'*

58. *L132: output => outputs*

59. *Caption Figure 2: bold => thick*

60. *L246: decrease => were lower; increasing => higher*

61. *L312: recognized as degrading process => recognized as a degrading process*

62. *L409: addition > additional*

References

Coleman, K. and Jenkinson, D. S.: RothC-26.3 - A Model for the turnover of carbon in soil, in: Evaluation of Soil Organic Matter Models, edited by: Powlson, D. S., Smith, P., and Smith, J. U., Springer Berlin Heidelberg, Berlin, Heidelberg, 237–246, https://doi.org/10.1007/978-3-642-61094-3_17, 1996.

European Commission: Regulation (EU) 2024/3012 of the European Parliament and of the Council of 27 November 2024 establishing a Union certification framework for permanent carbon removals, carbon farming and carbon storage in products, http://data.europa.eu/eli/reg/2024/3012/oj, Regulation (EU) No. 2024/3012, 2024.

Finke, P. a.: Modelling Soil Development Under Global Change, Springer Nature Switzerland, Cham, https://doi.org/10.1007/978-3-031-55583-1, 2024.

Finke, P. A., Samouëlian, A., Suarez-Bonnet, M., Laroche, B., and Cornu, S. S.: Assessing the usage potential of SoilGen2 to predict clay translocation under forest and agricultural land uses, European Journal of Soil Science, 66, 194–205, https://doi.org/10.1111/ejss.12190, 2015.

IUSS Working Group WRB: World reference base for soil resources 2022: International soil classification system for naming soils and creating legends for soil maps, 4th edition, International Union of Soil Sciences, Vienna, Austria, 2022.

Meng, X., Kooijman, A. M., Temme, A. J. A. M., and Cammeraat, E. L. H.: The current and future role of biota in soil-landscape evolution models, Earth-Science Reviews, 103945, https://doi.org/10.1016/j.earscirev.2022.103945, 2022.

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., O'Rourke, S., Richer-de-Forges, A. C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., Van Wesemael, B., and Winowiecki, L.: Soil carbon 4 per mille, Geoderma, 292, 59–86, https://doi.org/10.1016/j.geoderma.2017.01.002, 2017.

Ranathunga, K. N., Finke, P. A., Yin, Q., Verdoodt, A., and Yu, Y.: Soil modeling for soil loss tolerance estimations: Exploring natural baselines and long-term variations, Global and Planetary Change, 204, 103548, https://doi.org/10.1016/j.gloplacha.2021.103548, 2021.

Ranathunga, K. N., Finke, P. A., Yin, Q., and Yu, Y.: Calibrating SoilGen2 for interglacial soil evolution in the Chinese Loess Plateau considering soil parameters and the effect of dust addition rhythm, Quaternary International, 607, 100–112, https://doi.org/10.1016/j.quaint.2021.08.019, 2022.

Sallaba, F., Lehsten, D., Seaquist, J., and Sykes, M. T.: A rapid NPP meta-model for current and future climate and CO2 scenarios in Europe, Ecological Modelling, 302, 29–41, https://doi.org/10.1016/j.ecolmodel.2015.01.026, 2015.

Sauer, D., Finke, P., Sørensen, R., Sperstad, R., Schülli-Maurer, I., Høeg, H., and Stahr, K.: Testing a soil development model against southern Norway soil chronosequences, Quaternary International, 265, 18–31, https://doi.org/10.1016/j.quaint.2011.12.018, 2012.

Temme, A. J. A. M., Armitage, J., Attal, M., Gorp, W. van, Coulthard, T. J., and Schoorl, J. M.: Developing, choosing and using landscape evolution models to inform field-based landscape reconstruction studies, Earth Surface Processes and Landforms, 42, 2167–2183, https://doi.org/10.1002/esp.4162, 2017.

Van der Meij, W. M., Temme, A. J. A. M., Lin, H. S., Gerke, H. H., and Sommer, M.: On the role of hydrologic processes in soil and landscape evolution modeling: concepts, complications and partial solutions, Earth-Science Reviews, 185, 1088–1106, https://doi.org/10.1016/j.earscirev.2018.09.001, 2018.

Yu, Y. Y., Finke, P. A., Wu, H. B., and Guo, Z. T.: Sensitivity analysis and calibration of a soil carbon model (SoilGen2) in two contrasting loess forest soils, Geoscientific Model Development, 6, 29–44, https://doi.org/10.5194/gmd-6-29-2013, 2013.

Zhang, J., Quine, T. A., Ni, S., and Ge, F.: Stocks and dynamics of SOC in relation to soil redistribution by water and tillage erosion, Global Change Biology, 12, 1834–1841, https://doi.org/10.1111/j.1365-2486.2006.01206.x, 2006.