REVISION NOTES - Manuscript “A hybrid framework for the spin-up and initialization of distributed coupled ecohydrological-biogeochemical models” (GMD, egusphere-2025-4796) by Lian et al.

Editor

Thank you for the opportunity to revise our manuscript for potential publication in *Geoscientific Model Development*. In this second revision, we have addressed the remaining comments from Reviewer #1 and incorporated the final optional suggestions from Reviewer #2.

We also used this round of revision to correct minor typos throughout the manuscript. In addition, we would like to note a correction to the affiliation of one co-author, Athanasios Paschalis. In the previous version, two affiliations were listed (Imperial College London and University of Cyprus). This has now been corrected to retain only the University of Cyprus affiliation.

We would like to thank you and the reviewers for your time and insightful comments, which have significantly improved the quality of our work.

Reviewer #1

Thank you for the revised manuscript which has substantially improved, with concerns related to methodology, novelty adequately addressed. The time savings which have been practically realized are impressive. However, the revision falls short concerning the characterisation of the steady states and the applicability for coupled model simulations (section 3.4):

> **Reply:** We thank the Reviewer for the positive feedback on the revised manuscript and for the additional comments. In this second revision, we further clarified the criteria for defining steady-state conditions and showed that these conditions are reached in our analyses. We also expanded and clarified the comparison between the uncoupled and coupled steady states in Section 3.4.
>
> To facilitate the revision, we copied here extracts of modified text below the corresponding reply. For revisions in the manuscript, we use *Deleted:XX* **in gray for deleted text**, and *text* **in blue for added text**. Line numbers refer to the **revised manuscript version.**

(1) The chosen steady state criteria is orders of magnitude less compared to criteria applied in carbon cycle models (e.g. Friedlingstein et al 2024, 2025: https://doi.org/10.5194/essd-2025-659) which are < 1gC/m2/year not < 1gC/m2/day as used here. As a consequence the tolerance here is at least one order of magnitude higher than carbon cycle research. As nitrogen fluxes are usually one order of magnitude smaller than carbon fluxes and ideally the criteria reflects this. Consider adjusting the steady state criteria for carbon accordingly or discuss the implication of the criteria with respect to the signal size targeted in typical application of your type of model.

> **Reply:** We thank the Reviewer for this comment. One of the steady-state criteria used in this study was defined as an absolute linear trend (slope) smaller than 0.1 gC/m^2/day (or gN/m^2/day, line 246), which corresponds to 36.5 gC/m^2/year (or gN/m^2/year). This is indeed about one order of magnitude higher than the criterion suggested by the Reviewer.

Our choice to define the threshold based on daily values reflects the temporal resolution of the T&C model, which simulates carbon and nutrient dynamics at a daily timestep. As such, slopes were computed directly from the daily time series, without aggregating the outputs to annual values. Specifically, for **Fig. R1** (same as Fig. S3 in the Supplementary Information) and the steady-state diagnostics, we extracted the long-term trend component from each daily pool time series over the 9-year period and fitted a linear regression to this trend with time expressed in days. The resulting slope therefore has units of gC/m^2/day or gN/m^2/day.

Nevertheless, when converting the slopes shown in **Fig. R1** to annual units, the simulated trends also fall below 1 gC/m^2/day (or gN/m^2/day). For example, for the case of SOM POC we have $3.55 \cdot 10^{-6}$ gC/m^2/day · 365 days = 0.0013 gC/m^2/year, and the slope of SON is $0.0015 \cdot 10^{-6}$ gN/m^2/day, equivalent to 0.55 gN/m^2/year. This was the case across all 55 modeled pools, with the exception of the above-ground woody cellulose carbon pool and below-ground litter metabolic carbon pool. For the former, the daily slope is -0.007 gC/m^2/day, corresponding to -2.56 gC/m^2/year, thus exceeding the stricter annual criterion of 1 gC/m^2/year. However, it satisfies our third steady-state criterion, exhibiting a relative change of 0.7% between the first and last values of the fitted linear trend, which is below the 1% threshold. The latter pool has a slope of -1.07 gC/m^2/year, only slightly exceeding the strict threshold. In **Table R1** below we have reported the steady-state diagnostics for each biogeochemistry pool in a representative grid cell. The table includes daily and annual slopes, absolute and relative changes of the fitted trends, and the pass/fail evaluation for each applied criterion. Note that we considered steady state to be reached when all 55 pools met at least one of the three criteria.

We have added a sentence in line 248 to clarify the units used in our steady-state criterion and to note that the simulated steady-state trends are also within the limit discussed above. We also revised the caption of Fig. S3 to clarify the unit of the slopes.

"The steady state was considered reached when the long-term trends in all soil carbon and nutrient pools satisfied at least one of three predefined criteria over the final 9-year simulation period: (i) an absolute linear trend (slope) smaller than 0.1 [gC/m^2/day or gN/m^2/day], (ii) an absolute difference between the first and last values of the fitted linear trend smaller than 0.1 [gC/m^2 or gN/m^2], or (iii) a relative change between the first and last values of the fitted linear trend smaller than 1%. Note that the slope threshold of 0.1 [gC/m^2/day or gN/m^2/day] differs in units from those used in some previous studies (e.g. 1 gC/m^2/year in Friedlingstein et al., 2025), reflecting the daily temporal resolution of our model outputs. However, when converted to annual units, the simulated slopes are of comparable magnitude and generally remain within the range of this stricter criterion (see Deleted: an example in Fig. S3 for an example)."

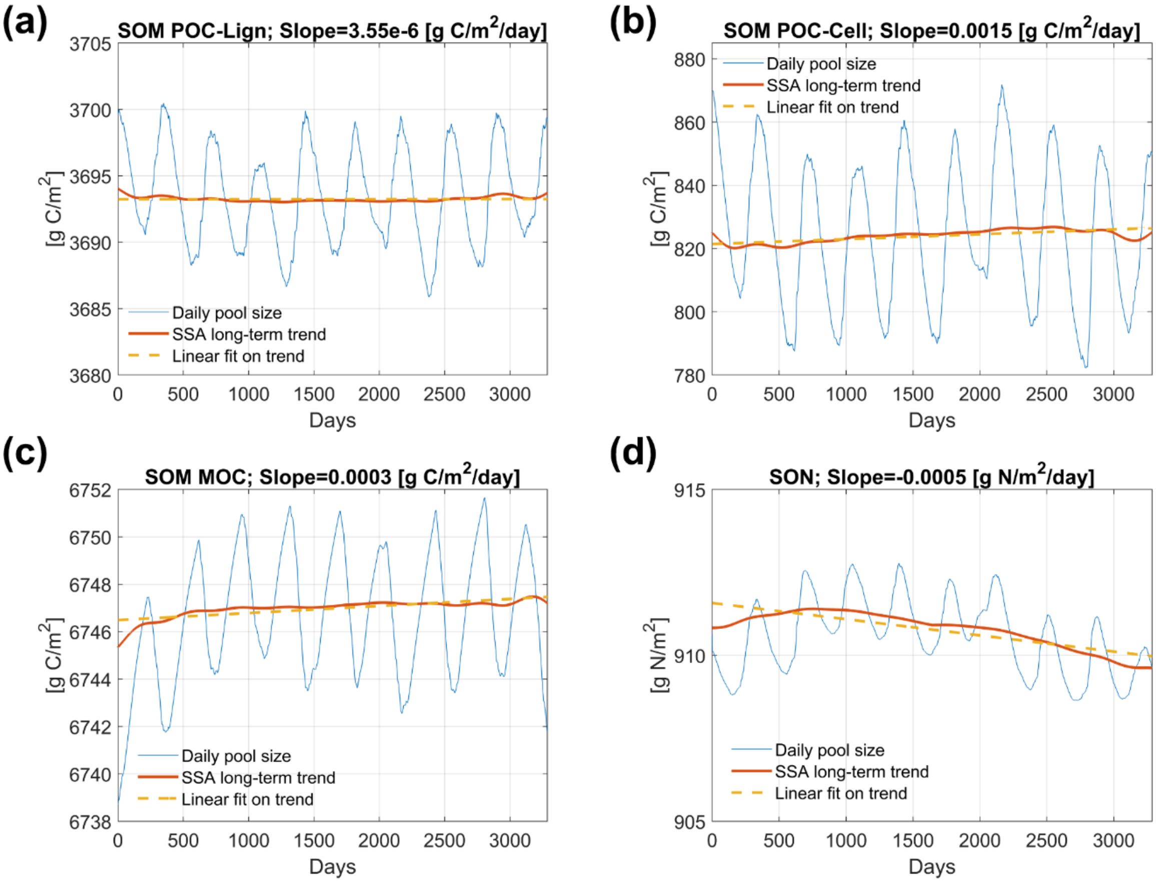

Fig. R1 (same as Fig. S3 in the supplementary information). *Time series of soil carbon and nitrogen pools during the 9-year coupled simulations for a representative grid cell. Long-term trends extracted using singular spectrum analysis (SSA), the corresponding linear fits, and the slopes (in $gC/m^2/day$ or $gN/m^2/day$) of the fitted trends are shown. Carbon pools include soil organic matter particulate organic carbon (SOM–POC) associated with lignin (SOM–POC–Lign; panel a) and cellulose/hemicellulose (SOM–POC–Cell; panel b), and mineral-associated organic carbon (SOM–MOC; panel c). Nitrogen pools include nitrogen in soil organic matter (SON; panel d). All pools are considered to have reached steady-state conditions.*

Table R1. *Steady-state diagnostics for soil carbon and nitrogen pools in the same representative grid cell shown in Fig. R1. The table lists, for each pool, the calculated values used to evaluate the steady-state criteria, including Slopedaily ($gX/m^2/day$), Slope_yearly ($gX/m^2/year$), the absolute difference between the first and last values of the fitted linear trend (gX/m^2), the relative change between the first and last values of the fitted linear trend (%), and the corresponding pass/fail classification (1/0) for each criterion. X can be carbon, nitrogen, phosphorus, or potassium.*

Pool	Slope_daily	Slope_yearly	Difference	Relative difference (%)	Pass i	Pass ii	Pass iii	Pass strict	SteadyState original	SteadyState strict
1	-6.50E-04	-2.37E-01	-2.13	4.0%	1	0	0	1	1	1
2	-8.18E-04	-2.99E-01	-2.69	1.5%	1	0	0	1	1	1
3	-2.15E-04	-7.85E-02	-0.71	2.4%	1	0	0	1	1	1
4	-7.00E-03	-2.56E+00	-23.04	0.7%	1	0	1	0	1	1
5	-2.30E-03	-8.49E-01	-7.64	0.7%	1	0	1	1	1	1
6	-2.90E-03	-1.07E+00	-9.62	15.1%	1	0	0	0	1	0
7	-1.70E-03	-6.34E-01	-5.71	1.5%	1	0	0	1	1	1
8	-1.80E-03	-6.48E-01	-5.83	6.4%	1	0	0	1	1	1
9	3.55E-06	1.30E-03	0.01	0.0%	1	1	1	1	1	1
10	1.50E-03	5.62E-01	5.05	0.6%	1	0	1	1	1	1
11	2.97E-04	1.09E-01	0.98	0.0%	1	0	1	1	1	1
12	2.91E-05	1.06E-02	0.10	1.0%	1	1	1	1	1	1
13	1.20E-04	4.36E-02	0.39	5.2%	1	0	0	1	1	1
14	1.67E-06	6.09E-04	0.01	8.1%	1	1	0	1	1	1
15	7.78E-07	2.84E-04	0.00	5.8%	1	1	0	1	1	1
16	6.96E-07	2.54E-04	0.00	8.1%	1	1	0	1	1	1
17	1.30E-06	4.73E-04	0.00	5.8%	1	1	0	1	1	1
18	6.03E-04	2.20E-01	1.98	6.1%	1	0	0	1	1	1
19	6.68E-04	2.44E-01	2.19	1.9%	1	0	0	1	1	1
20	-6.06E-04	-2.21E-01	-1.99	7.0%	1	0	0	1	1	1
21	8.58E-29	3.13E-26	0.00	0.0%	1	1	1	1	1	1
22	6.35E-05	2.32E-02	0.21	14.2%	1	0	0	1	1	1
23	-2.15E-04	-7.84E-02	-0.71	28.0%	1	0	0	1	1	1
24	-4.87E-05	-1.78E-02	-0.16	1.1%	1	0	0	1	1	1
25	-2.39E-04	-8.71E-02	-0.78	21.4%	1	0	0	1	1	1
26	-4.88E-04	-1.78E-01	-1.60	0.2%	1	0	1	1	1	1
27	1.16E-04	4.24E-02	0.38	6.1%	1	0	0	1	1	1
28	1.03E-04	3.77E-02	0.34	1.9%	1	0	0	1	1	1
29	-3.37E-05	-1.23E-02	-0.11	7.0%	1	0	0	1	1	1
30	8.58E-29	3.13E-26	0.00	0.0%	1	1	1	1	1	1
31	3.18E-06	1.20E-03	0.01	8.5%	1	1	0	1	1	1
32	-2.56E-07	-9.34E-05	0.00	0.8%	1	1	1	1	1	1
33	1.42E-06	5.17E-04	0.00	4.6%	1	1	0	1	1	1
34	6.35E-06	2.30E-03	0.02	14.2%	1	1	0	1	1	1
35	-1.11E-04	-4.03E-02	-0.36	59.8%	1	0	0	1	1	1
36	-3.48E-06	-1.30E-03	-0.01	1.1%	1	1	0	1	1	1
37	-1.14E-04	-4.17E-02	-0.37	55.1%	1	0	0	1	1	1
38	-2.74E-04	-1.00E-01	-0.90	0.5%	1	0	1	1	1	1
39	3.76E-05	1.37E-02	0.12	6.1%	1	0	0	1	1	1
40	1.68E-05	6.10E-03	0.06	1.9%	1	1	0	1	1	1
41	-5.05E-06	-1.80E-03	-0.02	7.0%	1	1	0	1	1	1
42	8.58E-29	3.13E-26	0.00	0.0%	1	1	1	1	1	1
43	5.99E-05	2.18E-02	0.20	0.5%	1	0	1	1	1	1
44	3.03E-05	1.11E-02	0.10	0.6%	1	1	1	1	1	1
45	-7.96E-04	-2.91E-01	-2.61	0.3%	1	0	1	1	1	1
46	3.17E-05	1.16E-02	0.10	2.7%	1	0	0	1	1	1
47	-1.15E-09	-4.21E-07	0.00	2.7%	1	1	0	1	1	1
48	-1.79E-05	-6.50E-03	-0.06	2.8%	1	1	0	1	1	1
49	-2.43E-05	-8.90E-03	-0.08	1.1%	1	1	0	1	1	1
50	-3.36E-05	-1.23E-02	-0.11	14.8%	1	0	0	1	1	1
51	1.01E-06	3.67E-04	0.00	0.1%	1	1	1	1	1	1
52	2.02E-05	7.40E-03	0.07	34.5%	1	1	0	1	1	1
53	1.99E-05	7.30E-03	0.07	36.1%	1	1	0	1	1	1
54	6.53E-06	0.0024	0.02	0.5%	1	1	1	1	1	1
55	3.11E-04	0.1135	1.02	0.3%	1	0	1	1	1	1

(2) Figure 7: the argument that the increase in SOC after switching from uncoupled to coupled simulation will approach the steady state of the coupled simulation is speculation. In a dynamic system this is not given.(For example, the progressive evolution of nutrient limitation with increasing nutrient immobilisation in accumulating SOC can reduce plant growth and subsequently soil C inputs which lead to a decline in SOC after an initial increase.). The authors could remove (or substantially revise)

section 3.4 which is not very convincing or provide proof that SOC in the model setups converge.

> **Reply:** We thank the Reviewer for this comment. In the present case, we confirmed that simulations starting either from the original initial condition (first column in **Fig. R2**, same as Fig. 7 in the manuscript) or from the uncoupled spin-up steady state (middle gray column) both converge to the same reference steady state (last column in Fig. R2). Consistent with our broader experience with this model, coupled simulations converge to the same steady state irrespective of the initial condition. To further support this point, we performed additional coupled T&C-BG simulations at one forest and one grassland site using two different initial conditions: (i) the original initial state and (ii) the steady state obtained from the uncoupled biogeochemistry-only spin-up. In both sites, simulations converged to the same long-term coupled reference steady state (**Fig. R3**).
>
> We agree with the Reviewer that, in many dynamic systems, the initial direction of change (e.g., increase or decrease in SOC) does not necessarily persist over time. Therefore, our previous wording - suggesting that the uncoupled spin-up steady state gradually moves toward the coupled steady state - may have been misleading. We have revised the relevant text (lines 373-375) to clarify that this behavior is specific to our simulations and may not occur in other cases.
>
> *Starting from this uncoupled spin-up steady state, an additional 27 years of simulation (3 × 9 years) were performed using the full T&C-BG model, with results shown in the three subsequent central columns. Such trajectories should not be interpreted as generally indicative of the direction toward the final steady state, as SOC and SON may follow non-monotonic pathways in dynamic systems. In the present simulations, however, both SOC and SON evolve toward the long-term coupled reference steady state during the first 27 years. The red box on the far right indicates the reference steady state achieved through long-term simulation …"*

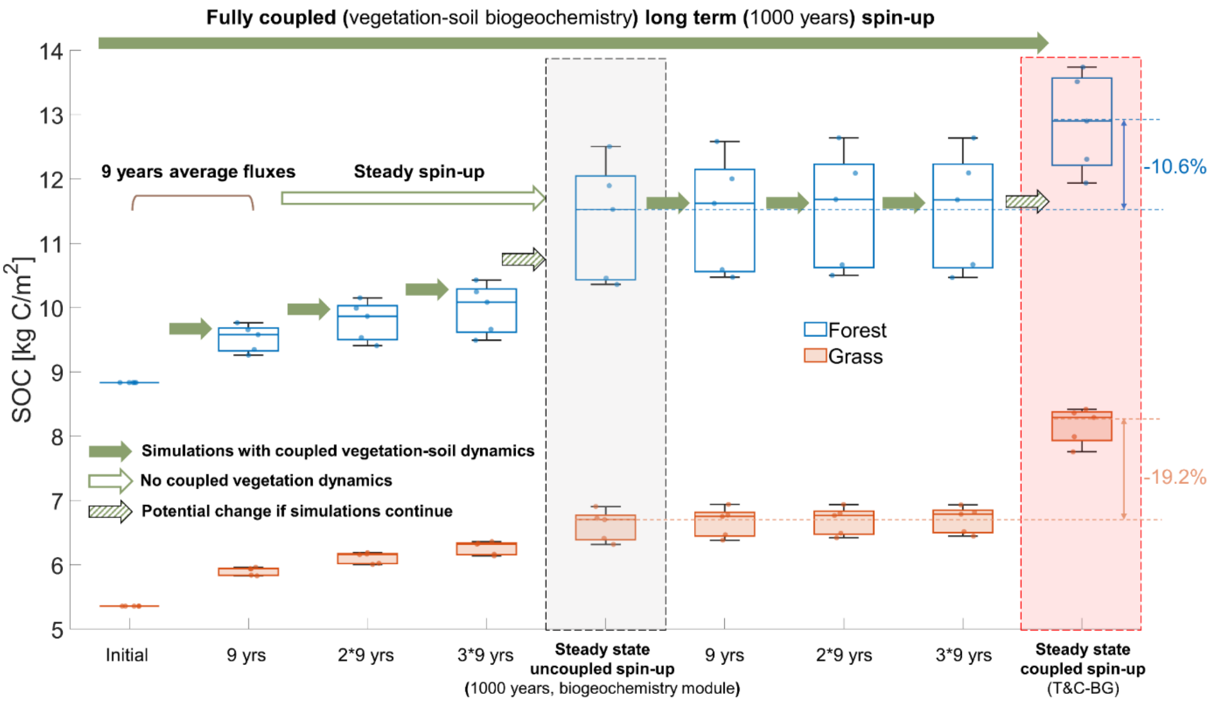

Fig. R2 (same as Fig. 7 in the manuscript). *Evolution of soil organic carbon (SOC) during different initialization schemes and simulation durations at 10 selected sites (Fig. 2h). Blue/orange boxplots represent forest/grassland sites. Arrows indicate simulation transitions* *between stages. The initial values (far left) are followed by 3*9-year simulations using the fully coupled T&C-BG model (solid green arrows). The average vegetation fluxes are extracted from the first 9 years to obtain a steady state condition with the biogeochemistry-only module (outlined green arrow, middle gray box). This is followed by three additional 9-year simulations. A comprehensive long-term spin-up using the fully coupled T&C-BG model (i.e., considering coupled vegetation and soil biogeochemical dynamics) is shown in the right red box. Striped arrows indicate the potential direction of SOC change if coupled simulations continue beyond the current duration.*

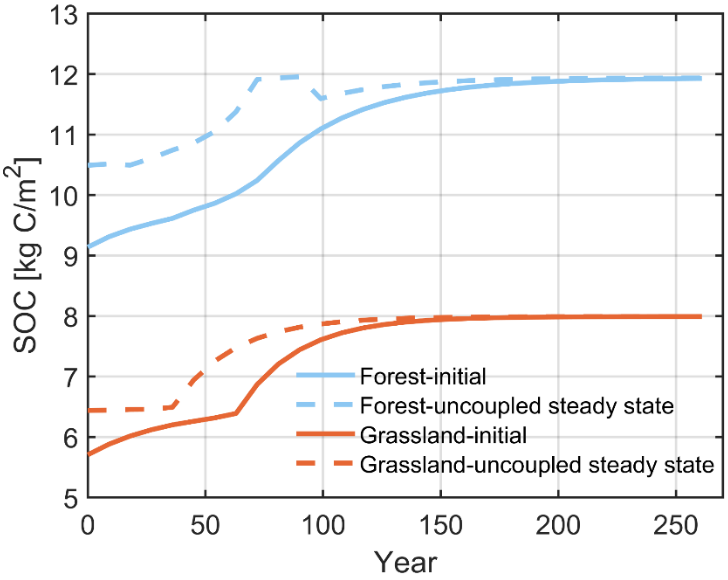

Fig. R3. *Convergence of soil organic carbon (SOC) in coupled T&C-BG simulations initialized from two alternative states at representative forest and grassland sites. The fully coupled model was run using either the original initial condition or the steady state obtained from the uncoupled biogeochemistry-only spin-up as the initial condition. At both sites, the SOC trajectories from these two initialization pathways converge to the same long-term coupled reference steady state, demonstrating convergence of SOC for the tested model setups.*

(3) The comparison of the difference between modelled states of up to 19.2% with uncertainties in SOC estimates is misleading (on Line 389c). The switch from uncoupled to coupled simulation (i.e. simulation setup) can cause soil carbon changes which are potentially of similar order of magnitude to the signals one wants to detect in the actual model experiments.

Reply: Thank you for this comment. We agree that the previous wording could have been misleading, as uncertainty in SOC measurements or estimates should not be directly compared with the bias introduced by the model setup. We have therefore removed this comparison from the manuscript and now only report the difference between the steady states obtained from the uncoupled and coupled spin-up approaches.

Line 383: *"A direct comparison of SOC between the uncoupled spin-up steady state (middle gray box in Fig. 7) and the reference steady state from the fully coupled plot-scale spin-up (red shading) reveals an average underestimation of 19.2% and 10.6% for grassland and forested sites, respectively (see also Table S3). Deleted: These values are within typically reported uncertainty ranges of SOC stock estimates (e.g., Fowler et al., 2023; Zhou et al., 2023; Even et al., 2025), and the The underestimation is further reduced for the case of SON (see Fig. S8)."*

(4) In addition, My concerns regarding the experimental design remain only partially addressed. The authors did not disentangle the effects of lateral transport from other sources of spatial heterogeneity which I think can be accepted, but the implications should be explicitly discussed.

> **Reply:** Thank you for this comment, please see our response below, together with our reply to comment (5).

Further comments:

(5) 'The experiment design does not allow to disentangle the effect of soil properties, climate, etc from the effect of vertical transport based on the experiments. As the deployment of RF from spinning up a model with vertical transport is the main novelty of this study I see that as a major shortcoming, and suggest an additional simulation is performed which differs from the benchmark case only from the omission of vertical Transport.' (reviewer's comment from revision 1)

- Variability in soil properties and climate are boundary conditions of the model as I understand, while the lateral(!) transport is within the model boundaries. Thus they are conceptually different. I still think demonstrating that effect of lateral transport would highlight the key novelty of this study, but

> **Reply:** We thank the Reviewer for this comment. We note that this comment is closely related to comment (4), and we therefore address them jointly. We agree that more explicitly illustrating the impact of lateral transport would be valuable.
>
> To isolate the effect of lateral transport, a relevant comparison is between simulated SOC and SON obtained from an initialization strategy that includes lateral transport and one that does not, while keeping other sources of spatial heterogeneity unchanged. Because performing separate 1D spin-ups for all 1859 cells in Erlenbach would be computationally expensive, we instead illustrate this effect using two representative cells. **Fig. R4** shows simulated SOC and SON in one forest cell and one grassland cell under different initialization strategies. Taking the spatial routing initialization as a benchmark, the bias caused by excluding lateral transport is 4.3% for forest SOC and 9.3% for grassland SOC, and 1.2% and 6.6% for forest and grassland SON, respectively. Note that these values are location-dependent and may vary with the topographic position of each cell, as the influence of lateral transport differs across the catchment. We also found that the homogeneous spin-up underestimates SOC and SON, whereas the spatial no-routing initialization overestimates them, indicating that different sources of spatial heterogeneity may affect initialization in different directions.
>
> At the same time, we emphasize that the objective of this study is not to disentangle the relative contributions of different sources of spatial heterogeneity, but to evaluate the proposed hybrid spin-up framework against more conventional initialization strategies. Such conventional approaches often represent landscape heterogeneity through grouped computational units, such as hydrological response units (HRUs), tiles, or other classes of cells sharing similar hydrological conditions (e.g., Chaney et al., 2018; Yang et al., 2025), which is conceptually similar to the vegetation/soil-combination-based 1D spin-up (Homogeneous spin up) considered here. We have revised lines 407-411 to better clarify these points.
>
> *"Overall, our experiments confirm the four hypotheses outlined in the introduction…Finally, an uncoupled biogeochemistry-only spin-up (H4) offers a computationally efficient first step, yielding steady states sufficiently close to the fully coupled solution, hence justifying its use in practical applications. Beyond the*

associated computational savings, this framework incorporates lateral transport between computational cells during spin-up, explicitly accounting for topography-induced lateral water redistribution. In this respect, it differs from more conventional grouped representations of landscape heterogeneity in hydrological and land surface models, such as hydrological response units (HRUs), tiles, or other classes of cells sharing similar characteristics (e.g., Chaney et al., 2018; Yang et al., 2025; Verhoef et al., 2026).”

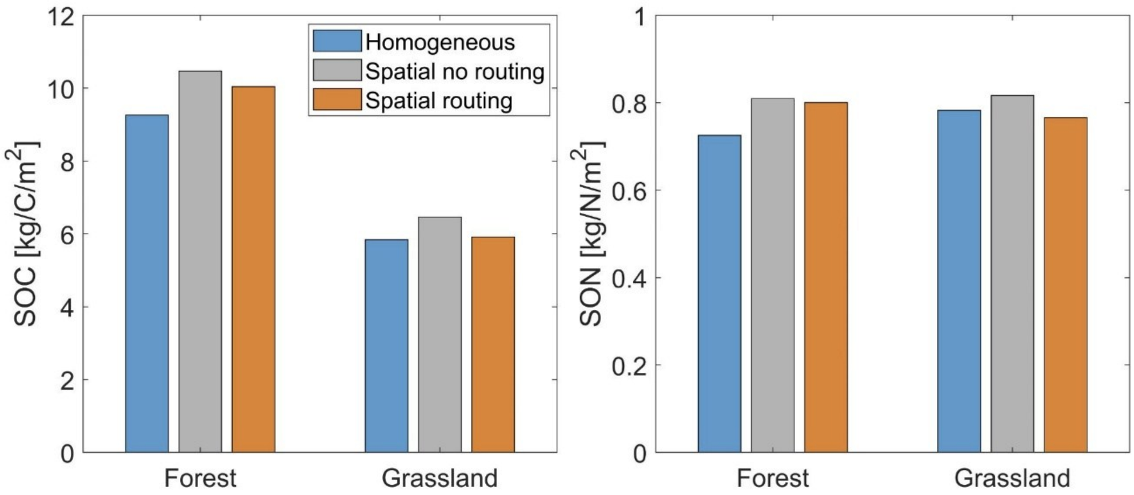

Fig. R4. *Simulated SOC and SON in a forest site and a grassland site after applying different initialization techniques. These techniques are: Homogeneous (the common practice of plot-scale spin-up for each soil–vegetation combination), Spatial no routing (plot-scale spin-up accounting for spatial heterogeneity in soil and meteorological conditions, but without lateral fluxes), and Spatial routing (consider spatial heterogeneity of soil, meteorological input, and lateral fluxes, extract from the benchmark 2D simulation results).*

(6)‘ In this study, given that predictor distributions are known and that the focus is on robustness across different sampling fractions rather than on optimal sampling design, we apply the transparent and reproducible stratified random approach.’ - This should be stated in the manuscript

<u>Reply:</u> Thank you for this suggestion. This was partly stated in the manuscript and we have now revised line 171 to further clarify this point.

“Because the focus of this study is on robustness across different sampling fractions rather than on optimal sampling design, we adopted a stratified random sampling approach based on vegetation types. Given that the distributions of topographic, soil, and vegetation characteristics across the full domain are known in advance and do not exhibit pronounced extreme tails, Deleted: stratified random sampling based on vegetation types this approach leads to training subsets that largely preserve these distributions and thus ensure representativeness (Fig. S2).”

(7) Interactions of P, K and vegetation in the model should be clarified in the manuscript (not only on the reply).

<u>Reply:</u> Thank you for this comment. We have revised the model description in lines 96-98 to clarify that vegetation growth in T&C-BG is regulated by the availability of soil N, P, and K.

"A soil biogeochemistry component, T&C-BG (Fatichi et al., 2019), was integrated…including nitrogen (N), phosphorus (P), and potassium (K). Vegetation growth is regulated by the availability of these nutrients, as plant tissue formation depends on flexible target stoichiometric constraints between carbon and nutrients, and nutrient limitation can constrain biomass production. The biogeochemistry module partitions plant litter into pools…"

(8) How were categorical variables treated in the RF? E.g did you use one-hot-encoding ? This comment was not addressed but should be clarified in the manuscript.

Reply: Thank you for this comment. We apologize for misunderstanding it in our previous revision. In most cases, vegetation type was represented by integer class labels and used directly as a predictor in the RF models, without applying one-hot encoding. For certain vegetation-specific pools, such as above-ground woody carbon, separate RF models were trained for grassland and forest cells, since these pools are only relevant to specific vegetation types. We have now clarified this in the Methods section.

Line 165: *"The covariates used as predictors include topographic features (elevation, slope, drainage area, and profile curvature), soil properties (sand content; clay content is additionally included in the random soil texture scenario described in the following section), and vegetation type (forest or grassland). In most RF models here, vegetation type was represented by integer class labels and used directly as a predictor. For vegetation-specific pools, such as above-ground woody carbon, separate RF models were trained for grassland and forest cells."*

(9) typo: 'the computation time for decoupled spin-up is negligible, see Table S6 despite the slight disagreement in steady states'

Reply: Thank you for catching this - We have now fixed the typo.

Line 388: "This provides a substantial gain in computational efficiency, as the computation time for decoupled spin-up is negligible Deleted:, see (Table S6) despite the slight disagreement in steady states, thus…"

Reference

Chaney, N. W., Van Huijgevoort, M. H., Shevliakova, E., Malyshev, S., Milly, P. C., Gauthier, P. P., & Sulman, B. N. (2018). Harnessing big data to rethink land heterogeneity in Earth system models. *Hydrology and Earth System Sciences*, *22*(6), 3311-3330.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., ... & Zeng, J. (2024). Global carbon budget 2024. Earth System Science Data Discussions, 2024, 1-133.

Yang, Y., Zhao, R., & Biswas, A. (2025). Delineating dynamic hydrological response units to improve simulations of extreme runoff events in changing environments. *Journal of Hydrology*, *656*, 133000.

Verhoef, A., Zeng, Y., Agam, N., Best, M., Bonetti, S., Boussetta, S., ... & Weihermüller, L. (2026). Rethinking soils in land surface models. Bulletin of the American Meteorological Society, 107(3), E665-E674.

Reviewer #2

The revised manuscript shows clear improvement and addresses the main concerns from my previous review. In particular, the authors have strengthened the machine learning component by adding cross-validation, more details on the random forest setup, and SHAP-based analysis. They also improved the description of the steady-state criteria and added explicit computational benchmarks. These changes make the work more transparent, reproducible, and convincing.

Some limitations remain, especially around the generalizability of the sampling strategy and the lack of testing in a different catchment. However, these points are now acknowledged and discussed. Overall, the manuscript is much stronger, and I think it is suitable for publication.

Below, you will find some minor comments that I leave to the discretion of the authors and editor to be implemented:

> **Reply:** We thank the Reviewer for the positive feedback. We are pleased that the revised version has addressed the concerns raised and that the manuscript is now considered suitable for publication. We also appreciate the additional minor comments provided and have carefully revised the manuscript accordingly, as detailed in our responses below.
>
> To facilitate the revision, we copied here extracts of modified text below the corresponding reply. For revisions in the manuscript, we use *Deleted:XX* **in gray for deleted text**, and *text* **in blue for added text**. Line numbers refer to the **revised manuscript version.**

1. The stratified sampling based on vegetation type improves reproducibility. It may be helpful to extend the discussion on how this approach could be applied to more heterogeneous catchments (e.g., by including elevation, soil type, or other gradients) and on how to assess whether the resulting sampling is sufficient.

> **Reply:** Thank you for your comment. We have extended the Discussion section to specify how to adjust and evaluate the sampling approach in more heterogeneous catchments.
>
> Line 457: *"...we recommend that, when implementing the proposed spin-up procedure in a new case study, the first step should be to assess the distribution of key environmental predictors in the study area (particularly vegetation type and soil texture) and to apply a proper sampling strategy to represent the predictor space. In more heterogeneous catchments, the sampling strategy could be extended by stratifying not only by vegetation type, but also by additional gradients such as elevation and soil type. The adequacy of the sampling can then be assessed by verifying whether the selected cells sufficiently cover the predictor space. It is also essential to evaluate whether the ecosystem is nutrient-limited…"*

2. The added cross-validation and SHAP analysis are strong improvements. A small clarification could help: even if the RF is not predicting an unknown reference state, it still performs spatial generalization, and this could be briefly acknowledged.

> **Reply:** Thank you for your suggestion, we have added this clarification in the Methods section.

> Line 127*: "The core strategy for the initialization framework proposed here involves two main steps: (1) performing a plot-scale spin-up in a limited number of selected cells while explicitly accounting for lateral fluxes of water, carbon, and nutrients, and (2) using a RF model to extrapolate the resulting steady-state pools across the entire model domain. While the RF is not used here to predict an independent external reference state, it enables spatial generalization from the sampled tracked cells to the remaining ones across the catchment.* A primitive version of this procedure……"

3. A final pass to simplify some sentences, especially in the Methods and Discussion section, could improve readability.

> **Reply:** Thank you very much. We have further streamlined several sentences in the Methods and especially in the Discussion to improve clarity and readability – some examples are reported below.
>
> Line 91: *"It includes modules for soil moisture dynamics and vegetation processes such as plant phenology, Deleted: dynamic carbon allocation Deleted: in multiple plant carbon pools, and tissue turnover (Fatichi et al., 2012b, 2016)."*
>
> Line 181: *"Second, RF-initialized simulations are compared with the benchmark to address H2, thus evaluating whether extrapolation from a subset of representative cells can reasonably approximate the benchmark Deleted: while reducing the computational requirements."*
>
> Line 197: *"The primary purpose of this analysis was pragmatic: to identify Deleted: evaluate which predictors Deleted: provide are essential Deleted: information for reproducing spatial variability, and which Deleted: variables could Deleted:potentially be omitted when reliable spatial data are unavailable. "*
>
> Line 402: *"Spatially explicit spin-up Deleted: (H1) is necessary to capture spatial variability in soil carbon and nutrient pools (H1), while RF-based extrapolation from a subset of representative cells Deleted: (H2) provides an efficient and accurate alternative to a fully distributed Deleted: and computationally expensive spin-up (H2). The required sampling fraction Deleted: (H3) is not universal (H3) but depends on catchment-specific heterogeneity and the choice of predictors. Finally, an uncoupled biogeochemistry-only spin-up Deleted: (H4) offers a computationally efficient first step (H4), yielding… "*
>
> Line 417: *"The Erlenbach catchment served Deleted: was used here as a testbed to evaluate the effectiveness of the new initialization scheme. Deleted: Results show that, for For this site, tracking only 40% of the cells Deleted: is was sufficient for the RF model to reconstruct over 90% of the spatial variability in key soil biogeochemical variables."*
>
> Line 422: *"Flatter topographies here did not necessarily reduce the required sample size, Deleted: both because the contributing area Deleted: in the two flatter topographies remains remained largely unchanged and Deleted: because heterogeneity of the domain (in soil and vegetation) soil–vegetation heterogeneity still needed to be sufficiently represented in the sampling."*
>
> Line 431: *"These findings align with previous studies showing that vegetation strongly influences the spatial distribution of SOC (e.g.,Kunkel et al., 2022; Yao et al., 2023) Deleted:, and that elevation is a dominant control Deleted: , at least in pronounced*

topographies such as Erlenbach because of the Deleted:, due to the elevation-related lapse rate control on air temperature (Stähli et al., 2021; Lian et al., 2025b). "

Line 442: *"This simplification may introduce a potential bias, Deleted: as different because soil textures can alter the degree of water limitation experienced by vegetation, Deleted: thereby affecting plant fluxes, and ultimately the initialization outcome."*

Line 469: *"To apply the proposed approach in systems that are not in steady state, Deleted: modifications are required. One a possible strategy Deleted: could be is to assume a pre-disturbance steady state..."*

Line 476: *"Deleted: It is important to note that In such cases, only the 1D spin-up component of the initialization scheme requires modification Deleted: in such cases; the RF algorithm remains applicable without change"*