Dear Taiqi Lian et al

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):

1. *(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.*
2. *(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.*
3. *(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. d*
4. *(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.*

Further comments:

'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.'

- *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*

‘ 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*

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

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.

typo: ‘the computation time for decoupled spin-up is negligible, see Table S6 despite the 395 slight disagreement in steady states’