Reply to RC1

I have read the manuscript describing soil informed neural networks to predict SOC density under sparse auxiliary data by leveraging multivariate learning and a soil-relation–informed ML.

The topic is timely for the DSM/ML community, and the paper's central idea, using constraints to improve plausibility and robustness under missing BD/CF and is potentially valuable.

However, there are several issues in the mathematical formulation, unit consistency, evaluation design, and soil-science framing that needs attention.

I outline some comments below.

RE: We thank the reviewer for the careful reading of the manuscript and for recognising the relevance and potential value of the study. We will address these points carefully and provide our responses in detail below.

-L.21 The Introduction implies SOC density as a DSM-driven stakeholder target. In practice: carbon accounting uses SOC stocks per area (depth-integrated), not only DSM.

RE: We agree that, in practice, carbon accounting is based on depth-integrated SOC stocks per unit area rather than SOC density alone. Our intention was not to present SOC density as the direct accounting variable, but to emphasize that SOC density is more closely related to SOC stocks than SOC content and is therefore a meaningful target in digital soil mapping. We will revise the Introduction to clarify this distinction.

- Depth handling (0–20 cm) and LUCAS 2018 exclusion, The study excluded 0–10 and 20–30 cm to focus on 0–20 cm. need to explain: whether 0–10 and 10–20 exist and why not used,

RE: This study was designed around the LUCAS topsoil layer at 0–20 cm, which is the standard sampling depth for the pre-2018 campaigns used here. In the 2018 campaign, a small subset of sites was additionally sampled by depth increments (0–10, 10–20, and 20–30 cm), but these represent only a minor fraction of the dataset (381 sites out of approximately 20,000). To maintain a consistent target depth across years, and because SOC is a relatively slow-changing property, we chose not to harmonize these few increment-based measurements into 0–20 cm, as such harmonization could introduce additional uncertainty and noise. We will clarify this in the revised manuscript.

- Unit consistency and dimensional correctness in SOC density (Eq. 1) with SOC content in g kg^{-1}, BD in g cm^{-3}, and output in kg m^{-3}.

While the formula can be numerically correct if the implicit conversions cancel , it is not dimensionally transparent and is easy to misapply. I suggest Rewrite Eq. (1) with explicit conversion constants, or (b) define the variables as mass fraction and kg m^{-3} explicitly.

RE: We thank the reviewer for this helpful suggestion. We will re-write Eq. (1) to include explicit conversion constants.

Also motivation says stakeholders care about "SOC density" more than content. In practice, stakeholders commonly want SOC as well as SOC stock per area (e.g., Mg ha^{-1} for a depth interval). Please temper the claim and clarify the end-use context (accounting, monitoring, agronomy, reporting).

RE: We agree that the original statement overstated the practical role of SOC density. Depending on the application, stakeholders may be interested in SOC content, SOC density, or SOC stock per unit area. We will therefore revise the Introduction to temper the claim and to clarify that carbon accounting and reporting typically rely on SOC stocks, whereas SOC density is discussed here as a relevant intermediate DSM target because of its closer relationship to stock estimation.

- SOC–BD mechanistic constraint (Eq. 2) and SOM=1.724 · SOC content

this is incorrect, because SOC content is in g kg^{-1} (not a fraction), so SOM becomes order 10–100+ (dimensionless), which makes (1−SOM) negative.

RE: We thank the reviewer for pointing this out. We agree that the original wording was not dimensionally correct as written. Since SOC content is expressed in g/kg, conversion to SOM must include scaling to a mass fraction before applying the factor 1.724. We will therefore correct the manuscript to make this explicit.

The Federer reference is not the origin of the equation. This equation of mixing is due to Adams WA. 1973. The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. J Soil Sci, 24 (1973), pp. 10-17

RE: We thank the reviewer for this correction. We agree that Adams (1973) should be cited as the original source of the mixing equation. The Federer reference was included because it provides additional theoretical and mechanistic explanation for the relationship. In the revised manuscript, we will cite Adams (1973) as the original reference and retain Federer et al. as supplementary theoretical support, with the wording adjusted accordingly.

- Coarse fragments. In Lucas, CF is measured in mass basis, how did you convert to volume basis as in Eq 2

RE: We agree that, in the original LUCAS dataset, coarse fragments are reported on a mass basis, whereas Eq. (2) requires volumetric coarse fragment content together with fine-earth bulk density. In this study, these variables were not derived by us from the raw LUCAS data, but adopted from the harmonized dataset of Pacini et al. (2023), as indicated in the manuscript. That dataset integrates the LUCAS 2018 bulk density campaign with coarse fragment information from earlier LUCAS surveys and provides both fine-earth bulk density and volumetric coarse fragment content for topsoil samples. We will add a brief clarification in the revised manuscript.

- L.85 The study promotes soil-informed ML, but at the same time uses 362 covariates from 15 groups, many of which are highly correlated, multi-scale, and partially redundant. That creates a tension between the stated philosophy and the modelling design.

RE: We thank the reviewer for this thoughtful comment. We agree that the use of a large covariate set creates a tension with a narrow interpretation of “soil-informed” modelling. In this study, however, “soil-informed” does not refer to restricting the model to a small or weakly correlated predictor set. Rather, it refers to incorporating soil knowledge into the model structure itself, by explicitly constraining the relationship between SOC and BD, while still allowing the model to learn from a broad set of environmental covariates known to influence SOC, bulk density and CF variation. The 362 covariates were not selected opportunistically, but assembled to represent established environmental controls on SOC across Europe, including climate, vegetation, terrain, land cover, soil moisture, and remote-sensing indicators . Importantly, the same covariate set was used for all three model architectures, so the comparison and the conclusions of the study concern the added value of the soil-informed model structure rather than the optimization of predictor selection.

- Cross-validation design likely suffers from leakage (repeated sites + spatial autocorrelation). It stated five-fold CV via random partitioning. But the dataset contains repeated measurements at the same sites across years. Random folds will almost certainly place the same site in both train and test folds (even if different years), inflating performance and plausibility diagnostics.

Additionally, DSM with dense covariates typically demands spatially blocked CV (or at least spatial buffering) to avoid optimistic estimates.

Use grouped CV by site ID so all time points for a site stay in one fold. Ideally, combine with spatial blocking (e.g., spatial k-fold) to reflect mapping/generalisation performance.

RE: We thank the reviewer for this important comment. We agree that, under observation-level random partitioning, repeated measurements from the same site across different survey years may in principle be assigned to different folds. However, we note that the direct accuracy assessment of SOC density is based only on the 2018 complete subset, as SOC density observations are only available for 2018 in the present dataset. Therefore, repeated-site leakage across years does not affect the main SOC density accuracy comparison.

For the multi-year temporal plausibility analysis, cross-validation predictions were not used, as SOC density observations are only available for 2018. Instead, the analysis was based solely on full-model predictions in order to assess temporal behaviour under a realistic application setting. For the temporal transferability test, the model was trained on the 2018 data and applied to the earlier survey years.

Moreover, we did not adopt spatial cross-validation because it is not necessarily the most appropriate approach for map accuracy assessment when the sampling design is not strongly clustered, as is the case for LUCAS. As discussed by de Bruin et al. (2022), even for clustered samples, spatial cross-validation is not always the most suitable assessment method, as it may yield overly pessimistic estimates of predictive performance.

- Targets are transformed to reduce skewness, and constrain to [0,1], achieved through "log transformation and scaling using a standard scaler." A standard scaler does not constrain to [0,1]; it standardises to mean 0, variance 1. Correct the description. Coul be Min–Max scaling?

RE: We thank the reviewer for pointing this out. The reviewer is correct that the term "standard scaler" was inaccurate. The target variables were not standardised to zero mean and unit variance, nor were they transformed using min–max scaling. Instead, we applied variable-specific transformations, consisting of logarithmic transformation where appropriate, followed by fixed multiplicative rescaling using predefined constants. We will correct the description in the revised manuscript accordingly.

- SOC density can only be "truly" validated where BD and CF exist, and here they exist only in 2018. The study claims robustness under sparse BD availability. But have not rigorously validated the "sparse BD reconstruction" claim. It just means

better internal consistency + smoother time series. But not neccessarily correct reconstruction when BD is truly absent.

RE: We thank the reviewer for this important comment. We agree that SOC density can only be directly validated where SOC, BD, and CF are jointly available, which in the present study is limited to the 2018 complete subset. We also agree that the current analyses do not provide a rigorous direct validation of SOC density reconstruction under truly absent BD observations. Our intention was not to claim that such reconstruction has been fully validated, but rather to show that, under sparse BD availability, the soil-informed model yields more internally consistent SOC density estimates, more plausible SOC–BD relationships, and smoother temporal behaviour than the purely data-driven alternatives. We will revise the manuscript to temper this claim and to state more explicitly that the evidence for robustness under sparse BD availability is indirect and based on plausibility and transferability diagnostics rather than direct multi-year SOC density validation.

- Temporal consistency filter appears logically inconsistent (likely a typo or mis-specified threshold). It wrote assume SOC changes < 0.5 g kg^{-1} yr^{-1}, but use a "conservative threshold of 50 g kg^{-1} yr^{-1} for the maximum absolute difference across measurements." This is confusing . Justify with citations and show sensitivity (how many series removed under alternative thresholds).

RE: We thank the reviewer for pointing this out. We agree that the original wording was confusing. The "50 g kg^{-1} yr^{-1}" should have been "50 g kg^{-1}" instead. And the value of 0.5 g/kg/yr, supported by the cited literature (Poeplau et al., 2011; Gubler et al., 2019), was intended only as an approximate reference for the expected rate of SOC change under normal conditions. In contrast, the threshold used in our data screening was a deliberately loose quality-control threshold intended to remove only clearly implausible repeated series, while avoiding excessive exclusion due to measurement uncertainty and other non-temporal sources of variation. We will revise the text to clarify this distinction and corrected the unit of the threshold accordingly.

- Some reported units and plausibility statements are incorrect. Example: "extreme changes exceeding 60 g cm^{-3}" for SOC density trajectories. SOC density is in kg m^{-3} (or equivalently g L^{-1}), not g cm^{-3}.

RE: We will correct this in the revised manuscript.

- Small sample sizes make stratified metrics unreliable (Table 3). In Table 3, Wetland has N = 2 but reports R^2 = 0.90. This is not meaningful.

RE: We agree that stratified performance metrics are not reliable for classes with extremely small sample sizes, and that values R2 = 0.9 for Wetland (N=2) should not be over-interpreted. We nevertheless chose to retain these results for completeness and transparency, so that readers can see the full stratified breakdown. In the revised manuscript, we will add a clear note to caution.

-MSE and R^2 are fine, but heavy tails and log transforms can distort interpretation. Add residual analysis (bias by SOC quantiles, BD quantiles)

RE: We thank the reviewer for this valuable suggestion. We will add residual analyses stratified by SOC and BD quantiles to improve the interpretation of model performance across their respective distributions.

For the joint SOC–BD space: consider distance metrics or coverage metrics i.e, how much predicted mass lies outside observed support.

RE: We thank the reviewer for this suggestion. We considered using distance- or coverage-based metrics to quantify how closely the predicted SOC–BD joint distributions reproduce the observed relationship. However, given the relatively limited number of paired SOC–BD observations, such metrics would be highly sensitive to how the observed support is defined, including choices related to binning, smoothing, or density thresholds. Moreover, the main differences among the models concern the overall shape and constraint of the joint distribution rather than a single scalar measure of distance or coverage. We therefore consider the direct comparison of the observed and predicted joint distributions to be more transparent and informative for the present analysis and will retain the current graphical assessment.

- Uncertainty quantification is missing, since "gains are modest", uncertainty reduction may be the key value proposition. Provide at least one uncertainty (ensembles, MC dropout, deep ensembles) for SOC density maps.

RE: We thank the reviewer for this valuable suggestion. We will add an uncertainty quantification analysis for the SOC density predictions to better assess model reliability and determine whether the soil-informed architecture provides benefits beyond the modest differences in predictive accuracy.

L.109 eq 2 is not mechanistic, still empirical

RE: We thank the reviewer for this comment. We agree that Eq. (2) should not be interpreted as a mechanistic model of soil processes in the broader sense, because it does not represent the causal dynamics of soil formation or SOC turnover. However, we do not consider it to be a purely empirical relationship either. The equation is based on a physical mixing formulation for organic and mineral soil components and has been

theoretically discussed and physically derived in previous studies, including Federer et al. (1993) and Robinson et al. (2022). We will therefore describe Eq. (2) more precisely as a physically motivated or semi-mechanistic mixing relationship, while avoiding referring to the overall SiNN framework as a mechanistic model.

Theory

A mass of "pure" mineral soil material (with no organic matter), m_m, occupies a volume, V_m, and therefore has a bulk density, D_{bm}, defined as m_m/V_m. Similarly, a mass of "pure" organic soil (with no mineral matter), m_o, occupies a volume, V_o, and has a bulk density, D_{bo}, defined as m_o/V_o. The F_o of some mixture of such mineral and organic masses is defined as

$$[1] \quad F_o = \frac{m_o}{(m_o + m_m)}$$

The mixture has a volume, V, and its D_b is $(m_o + m_m)/V$. We make the theoretical assumptions that (*i*) D_{bm} and D_{bo} remain constant in any mixture and (*ii*) the volume occupied by the organic fraction and the volume occupied by the mineral fraction in the mixture are additive, just as the masses are, so

$$[2] \quad V = V_m + V_o$$

Substituting the definitions of D_b, D_{bm}, D_{bo}, and F_o into eq. 2 leads to the theoretical equation

$$[3] \quad D_b = \frac{D_{bm}D_{bo}}{F_o D_{bm} + (1 - F_o)D_{bo}}$$

In a graph of D_b versus F_o, D_{bm} and D_{bo} are the D_b values at the end points $F_o = 0$ and $F_o = 1$, respectively. These two end points completely define the theoretical curve; its nonlinear shape derives from the underlying assumptions, not from the parameter values.