Reply to RC2

Overall assessment: The manuscript addresses an important challenge in DSM, i.e. estimating SOC density when bulk density and coarse fragment observations are scarce. The proposed SiNN is an interesting attempt to introduce soil knowledge into a ML framework through explicit SOC-BD relationships. The manuscript is generally well written and the idea is relevant to current efforts in physics-informed and process-informed ML for environmental sciences. However, the current manuscript does not yet fully demonstrate that the proposed architecture provides substantial advantages over simpler alternatives, nor does it fully establish the significance of the inferred latent parameters.

RE: We thank the reviewer for the constructive and detailed comments. We appreciate the reviewer's recognition of the relevance and potential value of introducing soil knowledge into machine-learning frameworks for digital soil mapping. In response to these comments, we will clarify the positioning of the proposed SiNN, moderate several claims, and better distinguish predictive accuracy from plausibility-oriented model evaluation. In the revised manuscript, we will clarify the terminology used for soil science-informed, process-informed, physics-informed, mechanistic, and constrained modelling approaches; provide a more explicit statement of the methodological contribution; expand the discussion of the plausibility and limitations of the inferred latent parameters; and add further analyses of model variability and uncertainty where appropriate.

Major comments

Introduction: Authors propose a soil science-informed neural network, but the distinction between "soil science-informed", "process-informed", "physics-informed", and "mechanistic" modelling remains unclear throughout the paper.

RE: We thank the reviewer for highlighting this lack of conceptual clarity. We agree that these terms should not be used interchangeably. In the revised manuscript, we will clarify their definitions and use them more consistently. Specifically, we will distinguish mechanistic or process-based models that explicitly represent causal soil processes, physics-informed models that incorporate physical laws or governing equations, physically constrained models that restrict predictions to physically admissible relationships, and soil science-informed models that embed established soil knowledge or soil-property relationships within a data-driven framework. Accordingly, we will revise the terminology throughout the manuscript and describe SiNN more precisely as a soil science-informed and physically constrained multitask neural network.

The proposed framework appears to introduce prior soil relationships through latent variables and constraints rather than through underlying physical equations.

RE: We agree with the reviewer's assessment. The proposed framework does not embed governing equations that describe dynamic soil processes. However, we would clarify that the embedded SOC–BD relationship, represented by the Federer–Adams mixing formulation is a physically motivated, semi-mechanistic representation of the contributions of organic and mineral components to bulk density. Within SiNN, this mixing model forms the final computational layer of the neural network, while the latent organic and mineral bulk-density parameters are learned from environmental covariates and allowed to vary spatially. The proposed approach can therefore be viewed as a hybrid of machine learning and semi-mechanistic modelling, following the broader concept of hybrid modelling described by Reichstein et al. (2019). In the revised manuscript, we will clarify this positioning and describe SiNN as a soil science-informed, physically constrained hybrid neural-network framework. We will also explain more explicitly that the latent variables are physically motivated parameters used to enforce plausible relationships among SOC content, bulk density, and SOC density.

I encourage the authors to more clearly position SiNN within the broader literature on process-informed and physics-informed ML and clarify whether the main contribution is the incorporation of a soil mixing relationship into a multitask NN architecture.

RE: We thank the reviewer for this helpful suggestion. We agree that the positioning of SiNN within the broader process-informed, physics-informed, and soil science-informed machine-learning literature should be made clearer. In the revised manuscript, we will highlight two related methodological contributions.

First, the multitask neural-network framework can jointly learn multiple soil properties while also using samples for which only a subset of the target variables is observed. This allows the model to exploit the abundant SOC content observations even where BD, CF, or SOC density measurements are unavailable, without requiring all target variables to be jointly observed at every training sample. Second, SiNN embeds an established, physically motivated SOC–BD mixing relationship and its associated latent organic and mineral bulk-density parameters within this multitask architecture. We will position this parameter-learning approach within the broader soil science-informed machine-learning literature, including related frameworks in which neural networks infer parameters of established soil relationships, such as soil-water retention functions. We will also clarify that this modelling strategy may be applicable to other soil-science problems in which physically meaningful parameters or relationships can be embedded within data-driven models.

Accordingly, we will describe SiNN as a soil science-informed, physically constrained hybrid multitask neural-network framework, rather than as a fully process-based or physics-informed model.

Authors should clarify (1) what makes SiNN fundamentally different from standard process/physics-informed options and (2) is the contribution primarily the embedding of a soil mixing equation?. Currently, the manuscript appears closer to a constrained multi-task NN than to a genuinely physics-informed framework.

RE: We thank the reviewer for this important comment. We agree that the distinction depends partly on how the Federer–Adams mixing formulation is characterised. In our view, this relationship is not merely an empirical constraint: it provides a physically motivated and semi-mechanistic representation of how organic and mineral components contribute to bulk density. Within SiNN, this formulation constitutes the final model layer, while the neural network learns the spatially varying latent parameters associated with organic and mineral bulk density. We will therefore clarify that SiNN is a hybrid, soil science-informed and process-informed neural-network framework, rather than simply a conventional multitask NN with an added admissibility constraint. Its main contribution is the integration of the Federer–Adams mixing formulation and its latent parameterisation into a multitask learning architecture. The multitask structure provides the additional benefit of learning jointly from samples with incomplete target measurements. At the same time, we will make clear that SiNN is not a complete dynamic process-based model of SOC turnover or soil formation.

It is difficult to determine whether the methodological innovation lies primarily in the latent parameterisation, the multitask learning strategy, or the incorporation of the SOC-BD relationship.

RE: We thank the reviewer for this comment. We agree that these components should be distinguished more clearly, although they are closely interconnected within the proposed framework. The multitask learning strategy provides the general architecture for jointly predicting multiple soil properties while allowing the model to learn from samples in which some target variables are missing. The defining contribution of SiNN is the incorporation of the SOC–BD mixing relationship into this architecture. The latent parameterisation of organic and mineral bulk density is an integral part of this formulation and provides additional physically interpretable outputs. Examining the magnitude and spatial distribution of these latent parameters also provides an additional plausibility check of the embedded hybrid model structure. In the revised manuscript, we will clarify the respective roles of these components and present the methodological contribution as their integrated combination rather than as three separate innovations.

Section 2.2: The manuscript includes 362 covariates from 15 groups. If the objective is to improve scientific interpretability through soil-informed modelling, why is the model driven by such a large and complex predictor space? Were there some form of importance analysis, feature attribution or group-level covariate importance. Otherwise, interpretability gains from latent soil parameters may be offset by opacity in the predictor space.

RE: We thank the reviewer for raising this important point. We agree that structural interpretability and predictor-level interpretability should be distinguished more clearly. In this study, the interpretability gain refers primarily to the model structure: the SOC–BD relationship is explicitly embedded in the architecture, and the latent oBD and mBD parameters can be examined for soil-science plausibility. It does not imply that the contribution of each of the 362 covariates is directly interpretable.

The broad covariate set was included to represent the diverse environmental controls on SOC, BD, and CF across Europe, including climate, topography, lithology, vegetation, land cover, soil moisture, and remote-sensing indicators. The same covariate set was used for UniNN, MultiNN, and SiNN so that differences among the models can be attributed to model architecture rather than predictor availability.

We did not perform feature-attribution or covariate-group importance analysis because the main objective of this study is to compare model architectures under a common predictor space. In the revised manuscript, we will clarify that the claimed interpretability gain is limited to the explicit soil-property relationships and latent parameterisation, while the high-dimensional predictor space remains difficult to interpret. We will also identify feature attribution and group-level covariate importance as relevant directions for future work.

Section 2.3: The current formulation appears closer to a multitask NN with embedded empirical constraints than to a genuinely process-informed or physics-informed framework.

RE: We agree with the reviewer’s assessment that the current framework is more appropriately described as a soil science-informed and physically constrained multitask neural network than as a fully process-informed or physics-informed model. However, we will clarify that the embedded constraint is not an arbitrary empirical relationship, but is based on an established SOC–BD mixing formulation and its associated latent organic and mineral bulk-density parameters. We will revise the terminology in Section 2.3 accordingly and avoid overstating the mechanistic nature of the framework.

Section 2.4: The manuscript reports performance metrics but does not provide information regarding fold-to-fold variability, confidence intervals, or statistical significance of differences among models. Given that many of the reported performance gains are relatively small, additional statistical evidence would help determine whether the observed differences exceed expected cross-validation variability.

RE: We thank the reviewer for raising this concern. We agree that the modest differences in R^2 and MSE do not provide sufficient evidence to conclude that SiNN is statistically superior to the simpler alternatives. However, this is not the main claim of the study. Rather, our results indicate that incorporating the SOC–BD constraint does not compromise predictive performance, while providing additional benefits in terms of physical plausibility, temporal coherence, and latent-parameter interpretation.

Given the current dataset, we consider it difficult to determine whether the small differences in point-estimate performance represent a reproducible gain. We therefore do not intend to interpret these differences as statistically conclusive. Instead, we will add an uncertainty quantification analysis for SOC density predictions and revise the manuscript to further moderate the comparison of predictive accuracy among the models. Where supported by the uncertainty analysis, we will also report uncertainty intervals for the relevant performance estimates or model differences.

Section 3: The reported improvements in SOC density prediction are marginal. The manuscript argues that plausibility and temporal consistency are the main benefits of SiNN, which may be true, but this shifts the contribution away from prediction accuracy. I recommend reportng explicitly on quantifying effect sizes of the improvements, significance of differences between models and variability across folds. Without those, it appears unclear whether the observed differences exceed normal cross-validation variability.

RE: We thank the reviewer for raising this important point. We agree that the differences in SOC density prediction accuracy among the models are marginal. We will therefore clarify that the manuscript does not claim that SiNN is statistically superior in terms of predictive accuracy. Rather, the main result is that incorporating the SOC–BD relationship maintains broadly comparable predictive performance while providing additional benefits in terms of structural plausibility, temporal coherence, and interpretation of the latent parameters.

To further assess model reliability, we will add an uncertainty quantification analysis for SOC density predictions. This will allow us to examine whether the physically constrained SiNN produces narrower or more stable predictive uncertainty than the purely data-driven alternatives. We will also revise the manuscript to avoid overstating the small differences in R^2 and MSE and to position predictive accuracy as a requirement that is maintained, rather than as the primary source of improvement.

Latent parameters remains speculative. A key claim is that latent parameters of oBD and mBD across land covers improve interpretability. However, interpretability is asserted rather than demonstrated. Do latent parameter estimates correspond to known soil processes and the estimated mBD values realistic across soil types? like the estimated oBD values, while assessing spatial coherency between them

RE: We thank the reviewer for this important comment. We agree that the interpretation of the latent oBD and mBD parameters should remain cautious. Because direct observations of these latent parameters are unavailable, the analyses are intended as plausibility assessments rather than independent validation.

In the revised manuscript, we will evaluate whether the inferred oBD and mBD values fall within literature-reported ranges and whether their variation across land-cover and soil-texture classes is physically reasonable. We will focus on land cover and texture because these variables are consistently available from the LUCAS survey, whereas a harmonised soil-type classification is not available for the present dataset and would require additional pedological information beyond texture alone.

We will also provide a qualitative assessment of the spatial plausibility and coherence of the inferred latent parameters based on their mapped patterns and their consistency with broad land-cover and soil-texture gradients. We will clarify that these analyses provide indirect evidence of plausibility rather than demonstrating that oBD and mBD uniquely or fully represent underlying soil processes. Accordingly, their interpretability will be described as structural and physically motivated rather than as independently validated pedological interpretation.

Section 3.3: SOC changes can be abrupt or gradual depending on land-use change or peatland disturbance. Smoother predictions could arise simply from stronger regularization. Authors should discuss this distinction more carefully

and avoid presenting smoothness as inherently desirable without validation against known temporal dynamics.

RE: We thank the reviewer for this helpful comment. We agree that smoother predictions should not be interpreted as inherently better, and this is not the intended interpretation of the temporal analysis. The purpose of this analysis is to assess whether the predicted temporal variation in SOC density contains extreme changes that are difficult to reconcile with plausible SOC dynamics over the LUCAS survey interval, rather than to validate the true temporal trajectory of SOC density.

In the revised manuscript, we will clarify this distinction and explicitly acknowledge that SOC changes may be abrupt following land-use change, peatland disturbance, or other major interventions. We will also note that smoother predictions may partly arise from the stronger structural constraints imposed by the SiNN architecture, rather than necessarily reflecting more accurate temporal dynamics. Accordingly, we will frame this analysis as an indirect plausibility assessment of extreme predicted changes, not as evidence that smoother trajectories are inherently more accurate. We will further clarify that direct validation of SOC density dynamics is not possible with the currently available data.

Section 4.2: It is argued that inferred oBD and mBD values provide additional insight into soil physical processes and improve interpretability. However, because no independent observations of these latent parameters are available, most interpretations remain hypothetical. The manuscript would be strengthened by demonstrating that estimated values fall within expected ranges across soil types and by providing additional evidence that the inferred spatial patterns correspond to known pedological gradients.

RE: We thank the reviewer for this important comment. We agree that, in the absence of independent observations, the interpretation of the inferred oBD and mBD parameters must remain cautious. The analyses of these latent parameters are intended as plausibility checks rather than as independent validation.

In the revised manuscript, we will assess whether the estimated oBD and mBD values fall within literature-reported ranges and whether their variation across land-cover classes and soil-texture space is physically reasonable. We will focus on land cover and soil texture because these variables are consistently available in the LUCAS dataset, whereas a harmonised soil-type classification is not available for the present analysis. We will also provide a qualitative assessment of whether the mapped latent-parameter patterns are spatially coherent and broadly consistent with known environmental and pedological gradients.

We will clarify that these comparisons provide only indirect evidence of plausibility. Accordingly, the interpretability of oBD and mBD will be described as structural and physically motivated, arising from the embedded SOC–BD relationship, rather than as definitive evidence that the latent parameters fully represent underlying pedological processes.

Section 4.3. The discussion links spatial patterns in latent parameters to land-cover and land-use effects. While these interpretations are plausible, they are currently not independently validated.

RE: We thank the reviewer for highlighting this point. We agree that the spatial correspondence between the inferred latent parameters and land-cover or land-use patterns does not constitute independent validation or causal attribution. In the revised manuscript, we will present these relationships more cautiously as plausible spatial associations and will clarify that independent observations or targeted field datasets would be required for rigorous validation.

Similar concerns apply to the subsequent derivation of porosity and compaction indicators. The manuscript presents these products as evidence of added soil-science value, yet they are derived from latent variables that themselves have not been independently validated.

RE: We thank the reviewer for this comment. We agree that the derived porosity and compaction indicators should be interpreted with the same caution as the latent oBD and mBD estimates from which they are calculated. Their purpose is to illustrate the type of additional soil-science information that may be derived from the SiNN parameterisation, rather than to present them as independently validated products or as evidence validating the latent parameters themselves. In addition, these indicators may serve as soil-physical consistency diagnostics. For example, observations or predictions that cannot be reconciled with established bulk-density–porosity relationships may indicate internally inconsistent or potentially problematic SOC or BD measurements. Nevertheless, because independent observations of oBD, mBD, and the derived indicators are unavailable, these analyses cannot constitute independent validation.

In the revised manuscript, we will describe these indicators as illustrative derived quantities and explicitly state that they inherit the assumptions and uncertainties associated with the inferred oBD and mBD values. We will also clarify that independent observations and targeted validation would be required before these indicators could be interpreted or used as standalone soil-physical products.

The manuscript states that latent parameter estimates enhance interpretability and improve transparency. This conclusion currently appears stronger than the presented evidence supports in the paper.

RE: We thank the reviewer for this important caution. We agree that the original wording may have overstated the degree of interpretability and transparency supported by the current evidence. Our intended claim is more limited: the latent parameterisation makes the embedded SOC–BD relationship and its organic and mineral bulk-density components more explicit within the model structure. This provides a form of structural interpretability, but does not demonstrate that the inferred latent parameters are independently validated representations of soil physical processes.

In the revised manuscript, we will therefore moderate these claims and state that the latent parameters may support model interpretation, conditional on the validity of the embedded SOC–BD relationship, the specified parameter constraints, and the plausibility of the inferred values. We will also revise the Conclusion to avoid implying that the SiNN framework is broadly transparent or that the latent parameters have been definitively validated.

Section 4.4: I appreciate that the authors acknowledge the lack of direct validation for temporal SOC density changes and discuss uncertainty quantification as future work. However, uncertainty may be particularly important in this study because the predictive gains achieved by SiNN are relatively modest. Demonstrating uncertainty reduction could potentially provide a stronger justification for the proposed framework than improvements in predictive accuracy alone.

RE: We thank the reviewer for this important suggestion. We agree that uncertainty is particularly relevant in this study because the differences in conventional predictive-accuracy metrics are modest. The primary motivation of the proposed framework is to embed established semi-mechanistic soil-science knowledge within a machine-learning architecture, thereby enabling the model to learn from sparse and partially observed soil datasets while maintaining physically meaningful relationships among SOC content, bulk density, coarse fragments, and SOC density. In the revised manuscript, we will therefore add an uncertainty quantification analysis for SOC density predictions to better characterise model reliability and compare the predictive uncertainty of SiNN with that of UniNN and MultiNN. This analysis will allow us to assess whether the embedded semi-mechanistic structure provides benefits through reduced or more stable uncertainty, even where improvements in point-estimate accuracy are limited. We will interpret the results cautiously and will not claim uncertainty reduction unless it is clearly supported by the analysis.

All three competing approaches are NN variants. This creates a somewhat narrow benchmark. For DSM applications, readers would likely expect comparison with other established methods such as tree-based options togther with NN or otherwise. If computational constraints prevent full benchmarking, the authors should justify why NN only comparisons are sufficient.

RE: We thank the reviewer for this comment. We agree that comparisons with established DSM methods such as Random Forests or Cubist would be valuable in a broader benchmarking study. However, the objective of the present study is not to identify the best-performing algorithm across different model families. Rather, it is to isolate the effects of three progressively structured modelling strategies within a common neural-network framework: independent single-target learning, multitask learning, and multitask learning with an embedded SOC–BD relationship and associated latent parameterisation.

Using neural networks for all three approaches ensures that the models have comparable representational capacity and training settings, so that observed differences can be attributed more directly to the multitask structure and soil-science constraint rather than to differences between fundamentally different algorithms. We will clarify this scope and rationale in the revised manuscript. We will also acknowledge that broader benchmarking against established tree-based DSM methods would be valuable future work, but lies beyond the focused methodological comparison of this study.

Throughout the manuscript, "soil science-informed", "physically constrained", and "mechanistic" appear to be used somewhat interchangeably. These concepts are not equivalent, mechanistic models represent causal processes, physical constraints enforce admissible solutions while soil-informed models may simply encode empirical relationships.

RE: We thank the reviewer for this clear and useful distinction. In the revised manuscript, we will revise the terminology throughout to avoid using these concepts interchangeably. However, we would also clarify that the embedded SOC–BD relationship is not merely an empirical constraint. Based on the physical reasoning presented by Federer et al. (1993) and Robinson et al. (2022), this SOC-BD formulation can be regarded as a physically motivated, semi-mechanistic mixing model that represents the contributions of organic and mineral components to bulk density.

In the revised manuscript, we will therefore describe SiNN as a soil science-informed hybrid neural-network framework that embeds a semi-mechanistic and physically constrained SOC–BD model. We will distinguish this from a fully mechanistic or process-based model of dynamic SOC change, since SiNN does not explicitly represent processes such as SOC turnover, transport, or soil formation. This terminology will preserve the mechanistic basis of the embedded mixing relationship while more accurately defining the scope of the overall framework.