Response to referee 2

We thank the editor and referee for their comments. Below, we respond to each comment (in blue text and preceded by **Authors:**).

Referee

Comment The manuscript is clearly written. Analyses are properly conducted. I have just a few questions. Since clay and silt content are also predicted from MIR spectra, how does the accuracy of these predictions influence the calculation of C_{def} and spectral modeling?

Authors: We thank the reviewer for this question. Although clay and silt contents used in the C_{def} calculation were estimated from MIR spectra, these estimates are accurate and comparable to standard hydrometer-based measurements. National MIR models for texture achieved strong performance (clay: $R^2 = 0.90$, CCC = 0.95, RMSE = 4.93%; silt: $R^2 = 0.84$, CCC = 0.92, RMSE = 6.97%), which is consistent with the accuracy reported for MIR-based particle-size prediction in other studies [Soriano-Disla et al., 2014, Hicks et al., 2015].

The use of MIR-predicted clay and silt, therefore, introduces an additional, but relatively small, source of uncertainty into C_{def} because any error in clay + silt propagates into the frontier-line–based estimate of MAOC capacity and hence into C_{def}. Given the low RMSE for clay and silt and the fact that C_{def} also depends on measured MAOC and on the frontier-line fit itself, we expect the contribution of MIR texture errors to the overall C_{def} uncertainty to be modest and not to materially affect the patterns or conclusions of either the frontier-line analysis or the subsequent spectral modelling, relative to using conventionally measured particle-size data (e.g. using the hydrometer method).

Comment The caption of Figure 3 – the last sentence is repeated.

Authors: Thank you for pointing this out. We will remove the repeated sentence in revision.

Comment What is the direct linear or nonlinear relationship between MAOC and C_{def}?

Authors: In our framework, C_{def} is not an independently varying quantity but is defined as the difference between the estimated maximum attainable MAOC (C_{Amax}) and the observed MAOC [Viscarra Rossel et al., 2024]. Because C_{def} is calculated directly from MAOC ($C_{def} = C_{Amax} - \text{MAOC}$), any apparent linear or nonlinear relationship between MAOC and C_{def} simply reflects this arithmetic definition rather than an additional ecological or mechanistic pattern, and therefore does not provide further insight.

Comment Since many spectral regions are identified that relate to organic groups, clay, and quartz, what is the model accuracy when using these soil properties to directly predict C_{def}?

Authors: We thank the reviewer for the pertinent question. We did not fit a separate model in which C_{def} is predicted solely from a restricted set of spectral regions associated with organic functional groups, clay minerals, and quartz. Instead, our aim was to exploit the full MIR spectrum, allowing the model to identify informative regions in a data-driven manner.

In MIR spectroscopy, many soil constituents share broad and overlapping absorption bands, and the same functional groups can occur in different molecular environments, causing shifts in peak positions and making precise a priori selection of 'pure' bands for specific components difficult. Spectral regions without distinct, easily assigned peaks can still carry useful contextual information when used in multivariate models. A priori, restricting the spectrum to hand-picked regions therefore risks excluding informative variance.

The CUBIST model employed here already performs implicit feature selection through its rule-based structure: only a subset of spectral variables enters the rules, while uninformative regions do not contribute to the prediction. As shown in our model interpretation, the learned rules concentrate on bands associated with organic matter and key minerals, so the effective predictor set is already focused on the most informative parts of the spectrum. For these reasons, we expect that an explicitly restricted-input model would not substantially improve predictive accuracy relative to the full-spectrum CUBIST model, and we have clarified this point in the revised manuscript.

Comment Since leave-site-out cross-validation is used in the study, how does the model accuracy compare when an independent validation is applied?

Authors: We thank the reviewer for this question and for catching the wording. In this study, we did not use leave-site-out cross-validation; instead, we implemented 10-fold cross-validation with grouping by site. This was a careless wording error on our part. We will correct this description in the revised manuscript.

To approximate an independent validation at the site level and to avoid data leakage, we used grouped 10-fold cross-validation, in which all samples from a given site are assigned to the same fold. In each iteration, the model is trained on $k-1$ folds and evaluated on the held-out fold, such that no site contributes samples to both training and validation within the same iteration.

Conceptually, both leave-site-out cross-validation and grouped k-fold cross-validation by site aim to assess model transferability to unseen sites by keeping sites intact. Leave-site-out CV holds out one site at a time and is attractive when the number of sites is modest, but can yield high-variance performance estimates when sites are numerous or highly heterogeneous. In contrast, a grouped 10-fold CV holds out a group of sites in each fold, providing a tractable number of resampling iterations and more stable estimates of model performance for larger (national) datasets. For these reasons, grouped k-fold cross-validation by site is a standard, widely accepted approach in soil modelling when (unbiased) external validation datasets are unavailable, and we consider it sufficient for assessing model robustness and generalisation across sites in this study. All performance metrics reported in the manuscript refer to this grouped cross-validation procedure.

Comment How does model performance vary with the three depth layers?

Authors: The table below shows the CUBIST model performance for C_{def} and MAOC for three depth layers, following the same modelling appraoch in the manuscript.

Depth	Variable	RMSE	R2	CCC
0-10 cm	mean C_{def}	4.397261	0.8649807	0.9234386
10-20 cm	mean C_{def}	2.670051	0.9179306	0.9513627
20 - 30 cm	mean C_{def}	3.542947	0.8770501	0.9271593
0-10 cm	MAOC	4.053939	0.7893657	0.8503351
10-20 cm	MAOC	1.549963	0.7948665	0.8636027
20 - 30 cm	MAOC	2.687933	0.8146391	0.8297

References

W Hicks, RA Viscarra Rossel, and S Tuomi. Developing the australian mid-infrared spectroscopic database using data from the australian soil resource information system. Soil Research, 53(8):922–931, 2015.

José M Soriano-Disla, Les J Janik, Raphael A Viscarra Rossel, Lynne M Macdonald, and Michael J McLaughlin. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Applied spectroscopy reviews, 49(2):139–186, 2014.

RA Viscarra Rossel, R Webster, M Zhang, Z Shen, K Dixon, Y-P Wang, and L Walden. How much organic carbon could the soil store? the carbon sequestration potential of australian soil. Global Change Biology, 30(1):e17053, 2024.