

REFEREE #2

Response: We sincerely thank the referee for the careful and constructive evaluation. We have addressed the manuscript issues directly and have expanded the analytical-uncertainty and local-model discussions. The most substantial new analysis is the explicit treatment of uncertainty introduced by calcimetric IC determination when DC-SOC is calculated as $TC - IC$.

Specific comments

Referee: "L105: Please correct the formatting of the CO₂ subscript."

Response: Thank you for detecting this formatting error. We have corrected CO₂ to CO₂ at the indicated location and will check the remainder of the manuscript for consistent chemical-formula formatting.

L104–108: This section appears thematically connected to the discussion of non-layer silicate minerals in L114. Consider grouping these sections together, either immediately before or after the section discussing layer-silicate clay minerals, to provide a more coherent progression of the discussion of mineralogical controls.

Response: The paragraph that was on L104-108 was awkward on its own. Although it connects to the subsequent review of mass loss from layer silicates, we thought lumping it into that paragraph would create a long, overly complicated paragraph. To improve readability, we added a topic sentence to this paragraph that better signals the grouping of paragraphs reviewing sources of mass loss other than SOM.

CHANGES IN THE MANUSCRIPT:

Introduction

Added the following sentence at the start of the two paragraphs reviewing sources of mass loss: "The causes of mass losses from sources other than SOM include a range of minerals that experience dehydration, dehydroxylation, or volatilization of CO₂ at temperatures that have been used for LOI."

Referee: "Figure 3: Please include explicit x- and y-axis labels indicating the variables being compared (i.e., predicted versus observed values). The figure should be interpretable independently of the main text."

Response: We revised Figure 3 to explicitly label the axes.

CHANGES IN THE MANUSCRIPT:

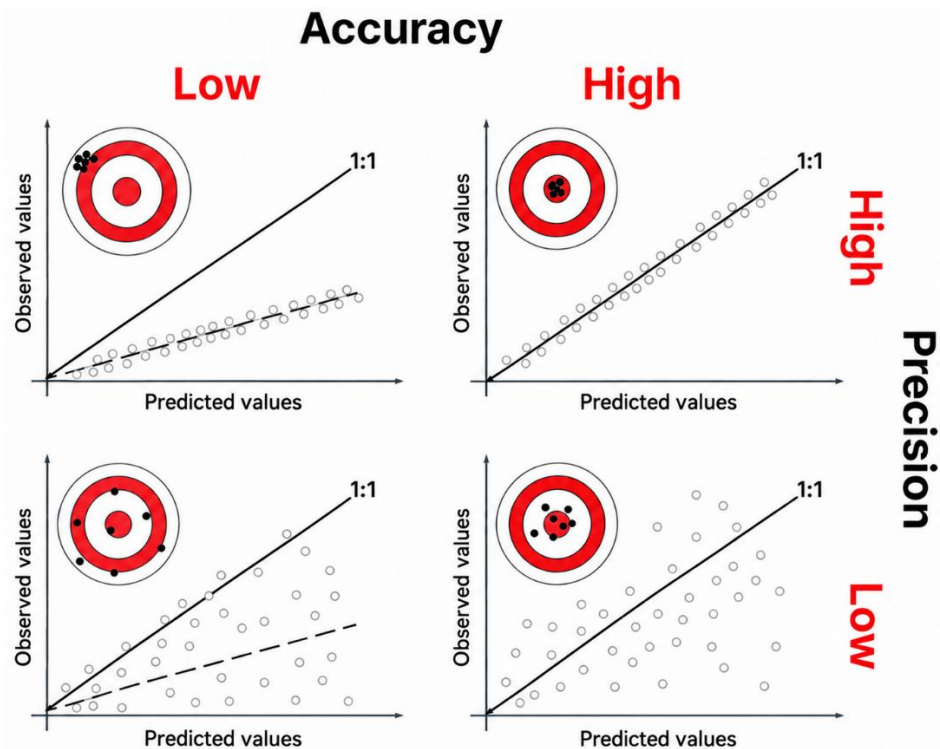


Figure 3: Conceptual illustrations of accuracy and precision, including the scatterplot scenario where the target is also a variable. Axis orientation follows Piñeiro et al. (2008), with predicted values shown on the x-axis and observed values on the y-axis. The 1:1 line represents the perfect match between predicted values and observed values. The upper-right panel demonstrates high accuracy and precision, with predictions closely matching observations and minimal scatter. The upper-left panel shows low accuracy but high precision, where predictions are consistent but systematically biased. The lower-right panel shows high accuracy but low precision, with predictions centered around the 1:1 line and exhibiting substantial scatter. The lower-left panel shows low accuracy and precision, with predictions that are both biased and highly variable.

Reference added:

Piñeiro, C., Perelman, S., Guerschman, J. P., and Paruelo, J. M.: How to evaluate models: Observed vs. predicted or predicted vs. observed?, *Ecol. Model.*, 216, 316-322, <https://doi.org/10.1016/j.ecolmodel.2008.05.006>, 2008.

Referee: “L334–335: The pattern described here could also be expressed in terms of positive and negative bias. Consider introducing this terminology, as it provides a standard and readily interpretable way of describing systematic deviations between predicted and observed values.”

Response: We agree that bias terminology improves interpretation. We have nevertheless avoided equating the regression intercept alone with overall prediction bias, because bias across the observed range depends on both slope and intercept. The revised wording uses positive/negative bias terminology while describing the intercept as one indicator of the systematic offset that the transfer equation compensates for.

CHANGES IN THE MANUSCRIPT:

Experimental design section 2.4

It is important to note that a positive intercept indicates that the method-transfer equation compensates for an apparent negative bias in LOI-SOM, which may arise from incomplete removal of organic matter during ignition. Conversely, a negative intercept indicates compensation for an apparent positive bias in LOI-SOM associated with additional non-SOM mass loss, such as dehydration, dehydroxylation, or carbonate-related mass loss (Pribyl, 2010). The intercept should therefore be interpreted as one indicator of systematic method-related bias rather than as a complete measure of prediction bias.

Referee: “L399–400: It may be useful to consider whether the uncertainty associated with the calcimetry (IC analysis) should also be incorporated by error propagation when assessing the variability of DC. This could potentially provide a more balanced estimate of the coefficient of variation associated with the DC measurements reported in Section 3.3.”

Response: We thank the referee for this important suggestion. In samples containing inorganic carbon, SOC was calculated as the difference between independently measured total C and IC ($SOC = TC - IC$). The uncertainty of the final DC-SOC value therefore includes contributions from both measurements.

We obtained the available calcimetry duplicate data from the laboratory. The IC dataset contained 25 duplicate pairs; 22 pairs had non-zero means and could be used to estimate a relative repeatability CV. The resulting IC repeatability CV was 2.51%. As an independent check, the 14 currently available TC duplicate pairs yielded a CV of 2.45%, close to the 2.60% DC repeatability originally reported from 73 replicate observations.

The original sample level TC and IC values corresponding specifically to those 73 replicate observations are no longer available, so we do not replace the empirical 2.60% in Table 4’s CV with a retrospectively reconstructed single value. Instead, we add a sensitivity analysis that propagates the TC and IC components and shows how the expected relative uncertainty in DC-SOC changes with the IC:SOC ratio. Assuming independent errors, $u(SOC) = \sqrt{(CV_{TC} \times TC)^2 + (CV_{IC} \times IC)^2}$

Using $CV_{TC} = 2.6\%$ and $CV_{IC} = 2.51\%$, the expected CV of DC-SOC is 2.6% when IC is absent, and increases to approximately 3.31%, 4.10%, and 5.77% when IC is 0.25, 0.50, and 1.0 times SOC, respectively. Thus, calcimetry contributes relatively little additional uncertainty when IC is small relative to SOC but becomes increasingly important when SOC is obtained as the difference between similar TC and IC concentrations.

CHANGES IN THE MANUSCRIPT:

Section 2.4 experimental design

For samples requiring correction for inorganic carbon, the contribution of calcimetric IC determination to analytical variability of DC-derived SOC was additionally evaluated by error propagation. Repeatability of IC determination was estimated from duplicate calcimetry measurements. Assuming independent analytical errors for TC and IC, the propagated standard uncertainty was calculated as:

$$u(SOC) = \sqrt{(CV_{TC} \times TC)^2 + (CV_{IC} \times IC)^2} \quad (11)$$

Where $u(SOC)$ is the propagated standard uncertainty of DC-derived SOC, expressed in percentage points of C; CV_{TC} and CV_{IC} are the coefficients of variation associated with TC determination by dry combustion and IC determination by calcimetry, respectively; and TC and IC are the measured total and inorganic carbon concentrations. While the corresponding relative uncertainty was calculated as:

$$CV_{SOC}(\%) = \frac{u(SOC)}{SOC} \times 100 \quad (12)$$

Because the original sample-level TC and IC values corresponding to the 73 DC replicate observations used to estimate the empirical DC repeatability CV were unavailable, propagated uncertainty could not be reconstructed retrospectively for those individual observations. Therefore, the effect of IC uncertainty was evaluated as a sensitivity analysis based on the ratio of IC to SOC:

$$CV_{SOC}(\%) = 100 \sqrt{\left[CV_{TC} \left(1 + \frac{IC}{SOC} \right) \right]^2 + \left[CV_{IC} \left(\frac{IC}{SOC} \right) \right]^2} \quad (13)$$

Where $\frac{IC}{SOC}$ represents the ratio of inorganic carbon to soil organic carbon, describing how the relative uncertainty of DC-derived SOC increases as IC becomes larger relative to SOC.

Section 3.3 – Results

Propagating the calcimetry analysis uncertainty showed that expected DC-SOC increases as IC becomes large relative to SOC: approximately 3.31%, 4.10%, and 5.77% for IC:SOC ratios of 0.25, 0.50, and 1.0, respectively. Therefore, the additional calcimetry contribution to uncertainty is small when IC is minor relative to SOC, but can become substantial when SOC is calculated as a small difference between TC and IC (Table 5).

Section 3.3. Repeatability and evaluation of measurement error

Table 5: Sensitivity analysis of propagated analytical variability in DC-SOC as a function of the IC-to-SOC ratio. Expected DC-SOC coefficients of variation were calculated assuming independent analytical errors for TC and IC, using an empirical TC/DC repeatability CV of 2.60% and an IC repeatability CV of 2.51% estimated from duplicate calcimetry measurements. Because SOC was calculated as TC – IC, the propagated relative uncertainty increases as IC becomes larger relative to SOC. Values represent expected analytical variability from error propagation rather than empirically observed CVs for the original 73 DC replicate measurements.

IC:SOC ratio	Interpretation	Expected DC-SOC CV(%)
0.00	No IC	2.60
0.25	IC= 25% of SOC	3.31
0.50	IC = 50% of SOC	4.10
1.00	IC = SOC	5.77

Referee: “L448: Please support this statement with an appropriate citation. In addition, it may be informative to express RMSE and MAE relative to the range of C concentrations represented in the dataset. Normalised RMSE and MAE would provide a relative measure of prediction error and may be more intuitively interpretable when assessing model performance across datasets or soil groups with different C ranges.”

Response: We agree with both points. First, the original statement implying that an error below approximately 0.5 percentage points of SOC generally has limited practical significance and was too categorical. Rather than defend a universal threshold, we will remove it and replace it with a context-dependent interpretation supported by literature on SOC change detection (Smith, 2004; Ellert et al., 2008). The practical significance of a given error depends on the SOC range, expected temporal change, monitoring objective, analytical uncertainty, and sampling design.

Second, we agree that normalized metrics provide useful context when comparing datasets or soil groups with different SOC ranges. We retained absolute RMSE and MAE as the primary accuracy metrics because they preserve the original SOC concentration units, and we added range-normalized RMSE and MAE as complementary measures. The definitions of these metrics were added to Section 2.3.3 (Accuracy) of the Materials and Methods, and the resulting normalized metrics are reported in Supplementary Table S2 and discussed in the Results and Discussion section

CHANGES IN THE MANUSCRIPT:

Material and Methods

Section 2.3.3

For comparisons among datasets or soil subsets with different SOC concentration ranges, RMSE and MAE were additionally normalized to the range of observed DC-SOC values. Range-normalized RMSE ($NRMSE_{range}$) and MAE ($NMAE_{range}$) were calculated as:

$$NRMSE_{range} = \frac{RMSE}{SOC_{max} - SOC_{min}} \times 100 \quad (7)$$

$$NMAE_{range} = \frac{MAE}{SOC_{max} - SOC_{min}} \times 100\% \quad (8)$$

Where SOC_{max} and SOC_{min} represent the maximum and minimum observed DC-SOC concentrations within the dataset or soil subset being evaluated. Absolute RMSE and MAE were retained as the primary error metrics because they preserve the original SOC concentration units, whereas the normalized metrics were used as complementary measures to facilitate comparisons among datasets with differing SOC ranges. Because range normalization is sensitive to the observed minimum and maximum values, normalized metrics were interpreted together with sample size and the underlying SOC distribution.

Results and discussion section 3.2.2. Effect of stratification by region

Although the localized method-transfer equations reduced error compared with the generalized method-transfer equation, the magnitude of these improvements remained modest. The maximum reductions were 0.46%, 0.32%, and 0.24% SOM for RMSE, MAE, and MedAE, respectively. However, these reductions were partly

influenced by differences in the SOC ranges represented by each landform region. Range normalization substantially changed the interpretation of some apparent improvements (Table S2). For the generalized dataset, $\text{NRMSE}_{\text{range}}$ values were 7.46%, 7.70%, and 9.46% for LOI360, LOI400, and LOI550, respectively. DML showed comparable normalized errors (7.78%, 7.85%, and 9.34%), whereas IES (10.17%, 10.02%, and 14.00%) and particularly NWP (16.74%, 16.92%, and 17.86%) showed greater relative errors. Although SIDP had substantially smaller absolute errors, its narrower SOC range reduced this apparent advantage after normalization with $\text{NRMSE}_{\text{range}}$ values of 6.58%, 11.16%, and 13.30% for LOI360, LOI400, and LOI550, respectively. Thus, only LOI360 retained a lower relative RMSE than the generalized method transfer equations.

These results demonstrate that lower absolute prediction errors within localized subsets do not necessarily indicate improved model performance when subsets span narrower SOC ranges. The practical importance of prediction error therefore depends on the magnitude and range of SOC being evaluated, together with analytical and spatial variability and the monitoring design (Smith, 2004; Ellert et al., 2008). Nevertheless, even minor improvements in accuracy may influence SOC detection over time, and the implications of these reduced errors for detecting SOC changes are examined further in Section 3.3.

References added

Smith, P.: How long before a change in soil organic carbon can be detected?, *Glob. Change Biol.*, 10, 1878-1883, [https://doi: 10.1111/j.1365-2486.2004.00854.x](https://doi:10.1111/j.1365-2486.2004.00854.x), 2004.

Ellert, B. H., Janzen, H. H., VandenBygaart, A. J., and Bremer, E.: Measuring change in soil organic carbon storage, in: *Soil Sampling and Methods of Analysis*, 2nd edn., edited by: Carter, M. R. and Gregorich, E. G., CRC Press, Boca Raton, FL, 25-38, 2008.

Supplementary Table S2 added:

Table S2. Absolute and range-normalized prediction errors for LOI method-transfer equations evaluated across the complete dataset and major soil subsets.

$\text{NRMSE}_{\text{range}}$ and $\text{NMAE}_{\text{range}}$ were calculated by dividing RMSE and MAE, respectively, by the range of observed DC-SOC concentrations within each evaluated subset and multiplying by 100. Normalized metrics are provided as complementary measures to facilitate comparison among datasets with different SOC ranges; absolute RMSE and MAE retain the original SOC concentration units. Range-normalized values for small subsets should be interpreted cautiously because the observed range is sensitive to sample size and extreme observations.

Dataset	LOI protocol	n	Observed DC-SOC range (%)	RMSE (% SOC)	MAE (% SOC)	NRMSE _{range} (%)	NMAE _{range} (%)	CCC
All	LOI360	423	0.00–7.58	0.566	0.361	7.46	4.76	0.801
All	LOI400	423	0.00–7.58	0.584	0.391	7.70	5.16	0.785
All	LOI550	423	0.00–7.58	0.717	0.527	9.46	6.95	0.636
With IC	LOI360	236	0.00–7.58	0.604	0.376	7.96	4.96	0.703
With IC	LOI400	236	0.00–7.58	0.633	0.419	8.35	5.52	0.664
With IC	LOI550	236	0.00–7.58	0.644	0.461	8.50	6.08	0.648
Without IC	LOI360	187	0.05–6.86	0.440	0.305	6.46	4.47	0.875
Without IC	LOI400	187	0.05–6.86	0.408	0.303	6.00	4.44	0.894
Without IC	LOI550	187	0.05–6.86	0.638	0.418	9.37	6.14	0.695
DML	LOI360	168	0.00–7.58	0.590	0.316	7.78	4.17	0.857
DML	LOI400	168	0.00–7.58	0.595	0.339	7.85	4.47	0.854
DML	LOI550	168	0.00–7.58	0.708	0.498	9.34	6.57	0.781
IES	LOI360	134	0.00–4.93	0.501	0.375	10.17	7.61	0.781
IES	LOI400	134	0.00–4.93	0.494	0.384	10.02	7.79	0.789
IES	LOI550	134	0.00–4.93	0.690	0.532	14.00	10.79	0.485
NWP	LOI360	92	0.00–3.56	0.596	0.443	16.74	12.45	0.569
NWP	LOI400	92	0.00–3.56	0.602	0.447	16.92	12.56	0.556
NWP	LOI550	92	0.00–3.56	0.636	0.479	17.86	13.45	0.478
SIDP	LOI360	29	0.16–2.08	0.126	0.106	6.58	5.53	0.976
SIDP	LOI400	29	0.16–2.08	0.214	0.171	11.16	8.91	0.927
SIDP	LOI550	29	0.16–2.08	0.255	0.211	13.30	11.00	0.892
A horizon	LOI360	203	0.07–7.58	0.713	0.476	9.50	6.34	0.688
A horizon	LOI400	203	0.07–7.58	0.709	0.456	9.44	6.07	0.694
A horizon	LOI550	203	0.07–7.58	0.836	0.554	11.13	7.37	0.516
B horizon	LOI360	152	0.00–2.63	0.336	0.215	12.79	8.16	0.492
B horizon	LOI400	152	0.00–2.63	0.344	0.230	13.09	8.73	0.454
B horizon	LOI550	152	0.00–2.63	0.368	0.266	14.00	10.13	0.323
C horizon	LOI360	68	0.00–5.63	0.399	0.288	7.09	5.12	0.825
C horizon	LOI400	68	0.00–5.63	0.490	0.354	8.71	6.29	0.710
C horizon	LOI550	68	0.00–5.63	0.520	0.389	9.24	6.91	0.661

Referee: “Table 3 and Figure 7: Please define all landform acronyms in the table/figure caption or in an accompanying footnote.”

Response: We added definitions for each landform-region acronym, so the table and figure can be interpreted without referring to the main text.

CHANGES IN THE MANUSCRIPT:

Table 3 and Figure 7 captions

DML = Des Moines Lobe; IES = Iowa Erosion Surface; NWP = Northwest Plain; SIDP = Southern Iowa Drift Plain.

Referee: “L338 and L488–522: The discussion of genetic horizon, clay content, and local models would benefit from further clarification. In L338, genetic horizon appears to have been used primarily as a proxy for clay mineralogy. However, clay content generally increases with depth, whereas substantial changes in clay mineralogy may occur only under particular pedogenic conditions. Consequently, L488’s suggestion that local models trained on sample subsets stratified according to clay content (and therefore potentially associated with depth or horizon) could improve the performance of LOI550 conversion models while maintaining parsimony is an interesting possibility. This is also consistent with the approach suggested in the recent literature (e.g. <https://doi.org/10.1016/j.geoderma.2025.117277>).

In contrast, the interpretation in L522 appears to somewhat discount this potential improvement. Please clarify the logical progression between these statements. Is the apparent limitation of local models primarily attributed to the relatively small sample sizes within the subsets, as suggested in L513–515? If so, this should be stated explicitly. Distinguishing between a lack of improvement due to insufficient sample size and a genuine lack of benefit from stratification would strengthen the interpretation and help readers understand the implications for future model development.”

Response: We thank the referee for this thoughtful comment and agree that the original manuscript did not distinguish clay content from clay mineralogy clearly enough. Soil horizon should not be interpreted as a direct proxy for clay mineralogy. Horizon differentiation can be associated with systematic changes in clay content, SOM concentration, carbonate occurrence, degree of weathering, and, in some pedogenic settings, mineralogical composition; however, increasing clay content with depth does not necessarily imply a corresponding change in clay mineralogy.

We also agree that the logical progression from the discussion of stratification's possible benefits to the later caution about local models was not sufficiently explicit. Our results do not demonstrate that stratification lacks value. Rather, they show two simultaneous patterns: localized models can improve performance when factors contributing to non-SOM mass loss are systematic within a subset, but increasingly specific stratification rapidly reduces calibration sample size. Several of the strongest apparent improvements were based on fewer than 10

observations, making it difficult to distinguish reproducible improvement from overfitting.

The limitation should therefore be attributed primarily to insufficient sample size within some subsets and incomplete characterization of the mechanistically relevant soil properties, not to an inherent lack of benefit from stratification.

LOI550 provides the clearest reason to explore clay-related predictors because it is most sensitive to mineral mass loss at high ignition temperature. We also appreciate the referee-provided reference to Richer-de-Forges et al. (2025). That paper concerns digital soil mapping of clay content rather than LOI methods transfer directly, so we use it only in the broader context of incorporating clay-related information and its uncertainty into predictive modelling.

CHANGES IN THE MANUSCRIPT:

Section 2.4

Landform region and soil horizon were evaluated as practical proxies for systematic differences in soil properties that could influence non-SOM mass loss during LOI measurements. Landform region primarily reflects contrasts in parent material and associated mineral assemblages, whereas horizon captures depth-related changes in properties such as clay content, SOM concentration, carbonate occurrence, weathering, and, in some soils, mineralogical composition. Soil horizon was therefore not assumed to represent clay mineralogy alone.

Section 3.2.3 rewritten version

In summarizing this analysis, we focus on CCC and prediction error because these metrics are most directly related to the practical performance of the method transfer equations. Across the three LOI protocols, localization produced apparent improvements in model performance for several of the evaluated subsets (Figures 8 and 9). Improvements were more frequent for LOI360 and LOI550, which were the protocols that tended to show greater systematic under- or overestimation relative to DC-SOC. This pattern suggests that localized method transfer equations can partly compensate for systematic interactions between analytical protocol and local soil characteristics. The largest apparent improvements were observed for LOI550, for which the higher ignition temperature increases the potential contribution of non-SOM mass loss from thermally sensitive soil minerals.

The influence of differing SOC ranges was also evident when models were stratified by soil horizon. Although B-horizon models had relatively small absolute RMSE values (0.34-0.37% SOC), their narrower observed SOC range (2.63% SOC) resulted in NRMSE_{range} values of 12.70-14.00%. These levels of error are greater than those obtained for the generalized dataset (7.46-9.46%). A horizon NRMSE_{range} values were also higher (9.44-11.13%), whereas C horizon values (7.09-9.24%)

were closer to those of the generalized dataset (Table S2), showing that apparent reductions in absolute error did not consistently translate into improved relative model performance across soil horizons.

These results should not be interpreted as evidence that stratification by soil properties cannot improve LOI method-transfer equations. Soil horizon is a relatively coarse stratification variable and should not be interpreted as a direct proxy for either clay content or clay mineralogy. Horizon differentiation may coincide with systematic changes in clay content, SOC concentration, carbonate occurrence, degree of weathering, and, under some pedogenic conditions, mineralogical composition, but these properties do not necessarily covary.

The strongest apparent improvement observed for LOI550 nevertheless provides a mechanistic reason to investigate clay-related predictors because this protocol is most susceptible to non-SOM mass loss from thermally sensitive minerals.

However, this would also be a reason to avoid LOI550 unless measuring at least clay content, if not clay mineralogy. Information about clay content could provide some information about clay-mineral abundance. Recent work in spatial soil prediction has likewise emphasized the importance of explicitly considering uncertainty associated with clay-related information when developing predictive models (Richer-de-Forges et al., 2025), although that study addressed digital soil mapping rather than LOI method transfer equations.

At the same time, increasingly specific stratification substantially reduced the number of observations available for model calibration. Several localized method transfer equations that appeared to outperform the generalized equations were based on fewer than 10 samples, increasing the risk of overfitting and making it difficult to determine whether those improvements would persist in a larger representation of the population.

Consequently, the present results cannot confidently distinguish between a genuine benefit of stratification for particular soil conditions and an apparent improvement arising from limited sample size. This interpretation is reinforced by cases in which CCC increased without a proportional decrease in MedAE. The limitation of the extremely localized models with small sample sizes evaluated here should therefore be attributed primarily to insufficient sample size within some subsets.

The contrast between SIDP and NWP illustrates these limitations. SIDP showed improvements in CCC for many of the evaluated segmentations, but the entire region contained only 29 samples, and progressively finer subdivisions contained substantially fewer observations. Moreover, although SIDP showed markedly smaller absolute prediction errors, normalization to its narrower SOC range substantially reduced these apparent advantages. Only LOI360 retained a lower

NRMSE_{range} than the generalized dataset, whereas normalized errors for LOI400 and LOI550 were higher (Table S2). In contrast, NWP showed little improvement from localization despite its larger sample size, suggesting that the factors contributing to non-SOM mass loss may be more heterogeneous or insufficiently represented by the available stratification variables.

Overall, the present dataset provides limited evidence that localization based solely on landform region, soil horizon, or IC occurrence produces a consistent and robust improvement over the generalized method transfer equations. This finding does not imply that localized models are inherently ineffective. Rather, reliable evaluation of localization requires sufficiently large calibration subsets and predictors that are mechanistically related to non-SOM mass loss. Quantitative clay content and mineralogical abundance are promising candidates for future evaluation, particularly for LOI550, but their potential benefits must be weighed against the costs of those additional lab measurements. Until such relationships are demonstrated using adequately sized and independently evaluated datasets, the generalized method transfer equations remain the more parsimonious option for broad application.

Reference suggested by referee added and included in the text:

Richer-de-Forges, A. C., Chen, S., Heuvelink, G. B. M., van der Westhuizen, S., Orton, T. G., Bourennane, H., & Arrouays, D. (2025). Does digital soil mapping prediction performance of soil texture improve when adding uncertain field texture estimates? A study based on clay content. *Geoderma*, 456, 117277. <https://doi.org/10.1016/j.geoderma.2025.117277>.

Referee: "Table 4 and figure captions throughout: Please ensure that all acronyms are defined within each table or figure caption."

Response: We agree. We reviewed every table and figure caption and defined acronyms at first occurrence within each caption so that each item can be interpreted independently of the main text.

CHANGES IN THE MANUSCRIPT:

For all tables and figure captions

All tables and figure captions were checked.

Referee: "Figure 10: Consider incorporating an indication of measurement uncertainty, for example standard deviation, assuming that the analytical uncertainty estimated in this study is appropriate for this purpose. It could also be valuable to illustrate the potential implications of this uncertainty in a practical context, such as an example of the associated trade-off or uncertainty in C credits."

This would further help demonstrate the practical significance of analytical uncertainty.”

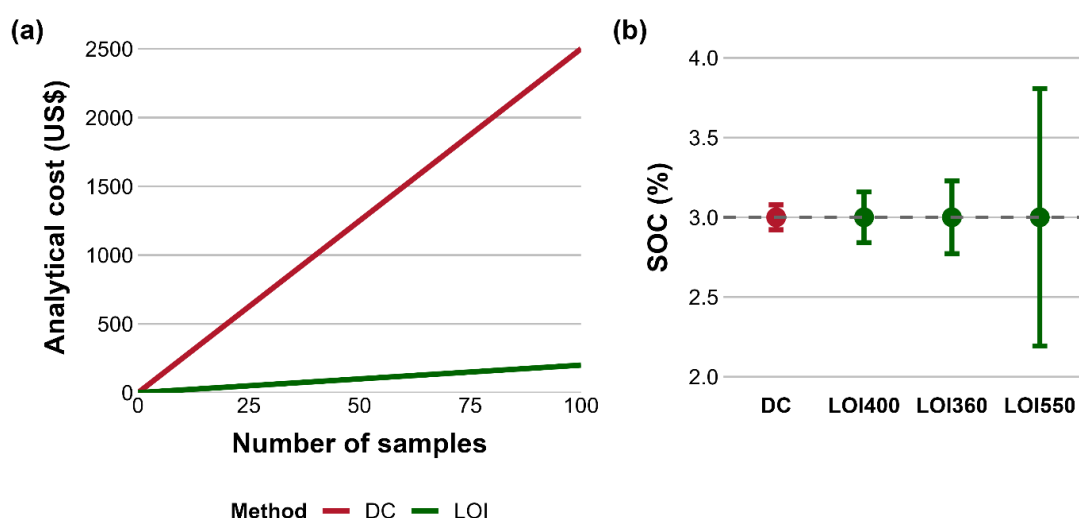
Response: We agree that Figure 10 should connect analytical cost with analytical repeatability more explicitly. Rather than adding error bars to cost bars, which could be misread as uncertainty in analytical price, we implemented a two-panel figure: panel (a) shows cumulative analytical cost as sample quantity increases, and panel (b) illustrates repeatability at a hypothetical SOC concentration of 3%. Using the observed CVs, ± 1 SD corresponds to approximately ± 0.078 percentage points SOC for DC, ± 0.159 for LOI400, ± 0.228 for LOI360, and ± 0.807 for LOI550.

We did not convert these analytical uncertainties directly into a monetary C-credit value because that calculation would require additional assumptions not evaluated here, including sampling depth, bulk density, area, stock-change design, verification rules, crediting methodology, and current C market prices. Instead, the revised text explains the practical implication that analytical uncertainty contributes to uncertainty in detected SOC change, while lower analytical cost can permit greater sampling density.

For consistency with the new IC uncertainty analysis, the DC error bar in the illustrative 3% SOC panel should be described explicitly as the empirical 2.60% repeatability estimate from Table 4. It does not include the scenario-dependent additional contribution from calcimetry in carbonate-containing samples; that effect is addressed separately in Section 3.3 through the IC:SOC sensitivity analysis.

CHANGES IN THE MANUSCRIPT:

Figure 10



Trade-off between analytical cost and repeatability for DC and LOI methods. (a) Estimated cumulative analytical cost as a function of sample quantity, assuming

costs of US\$25 per sample for DC and US\$2 per sample for LOI. The LOI cost applies to each evaluated LOI protocol. (b) Illustration of the analytical variability for a hypothetical sample containing 3% SOC. Points indicate the target SOC concentration (3%), and error bars represent $\pm$ one standard deviation estimated from the observed coefficients of variation for replicate analyses: 2.6% for DC, 5.3% for LOI400, 7.6% for LOI360, and 26.9% for LOI550. These correspond to approximate uncertainties of $\pm$ 0.08, $\pm$ 0.16, $\pm$ 0.23, and $\pm$ 0.81 percentage points, respectively. The dashed horizontal line indicates the target SOC concentration. Error bars illustrate analytical repeatability and should not be interpreted as confidence intervals. For carbonate-containing samples, the additional contribution of calcimetric IC determination to DC-derived SOC uncertainty is evaluated separately through error propagation in Section 3.3.

Section 4. Conclusions

Conclusions were reformulated in part based on the new additions (Black text represents the original conclusions, while blue text represent the reformulated part based on new additions).

This study shows that LOI should not be dismissed as an unreliable method for estimating SOC, but it should also not be [applied using simple proportional conversions that ignore systematic offsets between methods or by treating different LOI protocols as interchangeable](#). The results demonstrate that LOI400 (400 °C for 16 hours) stands out as a reliable alternative to DC for estimating SOC. While DC remains the benchmark for repeatability (analytical precision), its high cost and operational complexity restrict the number of samples that can be feasibly analyzed, which is a critical limitation when addressing spatial variability in SOC stocks. In addition, SOM is relevant to other environmental and agronomic management questions, thereby increasing incentives to monitor it beyond carbon markets alone.

LOI400 and LOI360 (360 °C for 2 hours) both deliver comparable precision and accuracy for predicting DC-SOC, with LOI400 exhibiting the highest repeatability among the LOI protocols. In contrast, LOI550 (550 °C for 3 hours) is less suitable for generalized SOC estimation due to increased measurement error resulting from mineral mass loss, particularly in soils with variable clay minerals, such as kaolinite, smectite, and illite. Although those additional sources of mass loss could be measured and compensated for in the method transfer equation, such additional analyses would defeat the purpose of keeping costs lower than those of DC.

[The method transfer equations evaluated in this research are relationships between LOI-SOM and DC-SOC measurements and should not be interpreted as estimates of the carbon concentration in SOM. Their purpose is to compensate empirically for systematic differences between analytical workflows. The](#)

evaluation of generalized versus localized equations showed that lower absolute prediction errors within localized subsets did not consistently translate into better relative performance when differences in SOC range were considered. In particular, range-normalized error metrics showed that some apparent improvements associated with landform region or soil horizon were partly attributable to narrower SOC ranges. Therefore, the present results should not be interpreted as evidence that stratification is ineffective, but rather that reliable localization requires sufficiently large calibration datasets and stratification variables that are mechanistically related to non-SOM mass loss.

Soil horizon and landform region were useful practical grouping variables but insufficient proxies for the soil properties driving differences in LOI performance. Quantitative clay content and, ideally, quantitative mineralogical information may provide more mechanistically relevant predictors, particularly for LOI550. However, several of the strongest apparent improvements in localized models occurred in subsets with fewer than 10 samples, limiting the ability to distinguish reproducible improvement from overfitting. Under the soil conditions represented in this study, generalized method transfer equations therefore remain the more parsimonious option for broad application, while localized equations warrant further evaluation when sufficiently large calibration datasets and relevant soil-property information are available.

Although the current literature does not indicate mass loss from carbonates at temperatures below 425 °C, excluding samples containing inorganic carbon substantially improved the precision of LOI-based SOC predictions and reduced their range-normalized prediction errors. In addition, the uncertainty analysis showed that DC-derived SOC in carbonate-containing samples includes uncertainty from both TC determination and the independent calcimetric determination of IC. This additional contribution is small when IC is minor relative to SOC but increases as IC becomes large relative to SOC. Thus, analytical uncertainty in the DC reference value itself should also be considered when interpreting LOI-DC agreement in carbonate-containing soils.

A key advantage of the LOI method is its low cost, which enables much higher sampling densities. This increased sampling capacity can offset the slightly higher measurement and prediction errors relative to DC, enabling more robust detection of spatial patterns and changes in SOC stocks. Moreover, LOI400 can also serve as a preliminary screening tool to identify zones of contrasting SOC and guide the selection of representative locations for subsequent DC analysis. In scenarios where resources are limited or spatial variability is high, LOI400 provides a practical solution for assessing SOC across large areas and verifying carbon markets. In addition, the ability to generate denser, lower-cost SOC data sets

supports emerging remote sensing and machine learning approaches that require extensive ground observations for model calibration and SOC digital mapping.

In summary, the choice between DC and LOI methods should be guided by the specific goals of SOC monitoring, available resources, and the degree of spatial variability in the study area. Among the three established LOI protocols evaluated in this study, LOI400 provided the best overall balance between predictive accuracy, repeatability, analytical cost, and scalability.

The generalized equation for converting LOI400-SOM to the DC-SOC equivalent was $\text{LOI-SOC} = 0.365 * \text{LOI-SOM} - 0.039$.