Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Texture-dependent bias in a European vis–NIR map of particulate and mineral-associated soil carbon

Md. Kamruzzaman

kamruzzaman.swe@gmail.com

Hungarian University of Agriculture and Life Sciences (MATE)

Short Report

Keywords: mineral-associated organic matter, particulate organic matter, carbon saturation, vis–NIR spectroscopy, soil texture, LUCAS

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11562265/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Particulate (POM) and mineral-associated organic matter (MAOM) carbon (C) predicted from visible–near-infrared (vis–NIR) spectra now inform a C saturation risk index for European agricultural soils. Their uncertainty was treated as random. I tested whether the error depends on soil texture at 6,399 LUCAS 2009 points not used for calibration, by comparing the summed predicted POM C and MAOM C with measured soil organic carbon (SOC). Overall, the sum tracked SOC (R^2 = 0.58; slope 0.92). At the same predicted value, however, measured SOC fell by 0.155 g kg^{-1} for every percent of sand (95% interval 0.138 to 0.174). Sandy soils were over-predicted by up to 14.0 g kg^{-1} and clayey soils under-predicted by up to 8.3 g kg^{-1}. In sandy soils the sum explained none of the variation in SOC (R^2 = − 0.31). A sand correction, validated by leaving out one 2° grid cell at a time, raised R^2 to 0.55 in sandy soils. Fractionation recovery did not explain the bias. The error is systematic and largest in coarse soils, where MAOM C capacity is smallest. It should be tested and corrected before saturation is computed from such maps.

1. Introduction

POM C is mainly plant-derived and cycles within years to decades, whereas MAOM C is bound to silt and clay and persists for decades to centuries (Lavallee et al., 2020). The amount of fine particles is thought to cap how much MAOM C a soil can hold (Hassink, 1997; Six et al., 2002; Georgiou et al., 2022), although no upper limit was detected in German agricultural soils (Begill et al., 2023).

Breure et al. (2025) calibrated a local partial least squares (PLS) model on 240 fractionated LUCAS 2009 samples (Lugato et al., 2021) and predicted POM C and MAOM C from vis–NIR spectra at 6,548 agricultural points. From MAOM C they derived the degree of saturation and a SOC risk index. They checked the summed fractions against measured SOC (R^2 = 0.59) and propagated uncertainty as an error normally distributed around each prediction, of one relative size for all soils.

Vis–NIR predictions of SOC are less accurate in sandy soils (Stevens et al., 2013), where the errors are dominated by over-estimation (Stenberg et al., 2010). If the fraction map inherits this, its error is a bias, not noise. I tested whether the error of the map depends on soil texture.

2. Materials and methods

Predicted POM C and MAOM C and the particle-size distribution of each point came from the Soil Carbon Risk Index dataset (Breure et al., 2025) of the European Soil Data Centre (Panagos et al., 2022), and measured SOC (ISO 10694) from the LUCAS 2009 topsoil data (Tóth et al., 2013). In all, 6,399 of the 6,548 predictions had a measured value, in 22 countries. None is among the 330 samples with measured fractions.

No point holds both a predicted and a measured fraction, so I evaluated the sum, as Breure et al. (2025) did; a difference shows that at least one fraction is wrong. I regressed measured on predicted SOC, because the reverse regression adds a spurious slope (Piñeiro et al., 2008), and summarized bias within quintiles of the prediction. Points were grouped as clayey (clay ≥ 35%), sandy (sand ≥ 65%), silty (silt ≥ 50%) or medium, in that order. Sand or clay content was added to the regression to test whether texture explains SOC beyond the prediction. Intervals are 95% percentile intervals from 2,000 bootstrap resamples of the 181 grid cells of 2° that

hold the points (Field and Welsh, 2007). A correction that regressed SOC on the prediction, sand content and their product was validated by leaving out one cell at a time. Because size fractionation does not always recover all SOC (Poeplau et al., 2018), I computed recovery, (POM C + MAOM C)/SOC, by texture class in the 235 cropland and grassland samples with measured fractions (Lugato et al., 2021).

3. Results and discussion

Overall, the sum tracked SOC with R^2 = 0.58, a bias of + 0.69 g kg^{-1} and a median absolute percentage error (MdAPE) of 24.7% (Table 1), close to the R^2 = 0.59 of Breure et al. (2025). The slope of measured on predicted SOC was 0.92 (0.88 to 0.97).

Table 1
Accuracy of the summed predicted POM C and MAOM C against measured SOC at 6,399 LUCAS 2009 points, as published and after a texture correction. The correction regresses measured SOC on the prediction, sand content and their product, and each point is predicted by a model fitted without its 2° grid cell. Bias is predicted minus measured SOC in g kg^{-1}; median SOC is in g kg^{-1}. Texture classes: clayey, clay ≥ 35%; sandy, sand ≥ 65%; silty, silt ≥ 50%; medium, the remainder. POM, particulate organic matter; MAOM, mineral-associated organic matter; C, carbon; SOC, soil organic carbon; MdAPE, median absolute percentage error.

Texture class	*n*	Median SOC	Bias, published	MdAPE (%), published	R^2, published	Bias, corrected	MdAPE (%), corrected	R^2, corrected
All points	6,399	17.3	0.69	24.7	0.58	0.01	21.8	0.69
Sandy	1,040	11.4	6.57	50.2	−0.31	0.22	25.0	0.55
Medium	2,754	18.3	1.85	25.8	0.59	−0.08	23.2	0.65
Silty	1,913	17.1	−2.31	19.5	0.59	0.13	20.2	0.64
Clayey	692	26.8	−4.50	19.2	0.63	−0.30	18.2	0.73

Within texture classes the map was biased (Fig. 1a). At the same predicted value, measured SOC fell by 0.155 g kg^{-1} per percent of sand (0.138 to 0.174) and rose by 0.295 g kg^{-1} per percent of clay (0.263 to 0.329). The departure from the overall line fell steadily from + 5.1 g kg^{-1} below 10% sand to − 7.6 g kg^{-1} above 80% (Fig. 1b). Sandy soils were over-predicted in every quintile of the prediction, by 0.9 to 14.0 g kg^{-1}. Clayey soils were under-predicted by 2.3 to 8.3 g kg^{-1}. In sandy soils the prediction explained none of the variation in SOC (R^2 = − 0.31; MdAPE 50.2%).

Recovery does not explain this. Its median was 0.98 in sandy (*n* = 35) and 1.03 in clayey soils (*n* = 48), and recovery above one in clayey soils would push the sum above SOC, not below it. The pattern matches the behavior of vis–NIR spectra in sandy soils, where quartz scattering may mask organic absorption and organic matter may dominate the spectrum of the sand matrix (Stenberg et al., 2010); adding sand as a predictor improved LUCAS SOC predictions in sandy soils (Stevens et al., 2013). Because texture is measured at every point, the bias can be corrected, and the authors of the map have since suggested testing covariates such as sand content in local PLS (Breure et al., 2026). A sand term fitted without the cell being predicted raised R^2 in sandy soils to 0.55 and halved their MdAPE to 25.0% (Table 1).

Which fraction carries the excess cannot be told from the sum. If it is MAOM C, the saturation of sandy soils is overestimated, because a small capacity turns a small absolute error into a large relative one. Breure et al. (2025) found MAOM C above the estimated capacity in sandy soils, which they attributed to the boundary-line regressions, and placed most high-risk land around the Baltic Sea, in northern Germany and in eastern England. Fine POM that passes into the MAOM fraction can also inflate MAOM C in coarse soils (Champiny et al., 2026). Texture-dependent prediction bias is a third possible cause; prediction bias has been proposed before to explain an apparent MAOM C limit in LUCAS data (Cotrufo et al., 2019; Begill et al., 2023). An uncertainty analysis centered on the predictions cannot represent such a bias. The test covers the sum only, and the points are the applicability-domain subset of one survey year, not a probability sample (Brus et al., 2011).

4. Conclusions

The vis–NIR map of POM C and MAOM C is close to unbiased over all European agricultural soils but biased within texture classes. Before MAOM C saturation is computed from such maps, the error should be tested against texture, corrected where texture is known, and its uncertainty conditioned on texture.

Declarations

CRediT authorship contribution statement

Md. Kamruzzaman: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI in the writing process

The author conceived and designed the study, formulated the research question, and determined the data selection and analytical approach, retaining full responsibility for the scientific content and conclusions. Generative AI was used as a supporting tool to revise the analysis code, to edit manuscript language and readability and to review for internal consistency and author independently verified. To support transparency and reproducibility, the stored summary data underlying the reported figures and the scripts required to reproduce them are provided as supplementary material.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics statement

This study is a secondary analysis of publicly available, previously published datasets and involved no new sampling of soil, and no human or animal subjects. Ethical approval was therefore not applicable.

Data availability

The Soil Carbon Risk Index dataset, the LUCAS 2009 topsoil data and the soil organic matter fractions dataset are distributed by the European Soil Data Centre (ESDAC) under standard registration. Their license does not allow them to be redistributed, so they are not included. The supplementary archive contains the analysis script, the R script that draws Figure 1 and Table 1, and the stored summary tables.

Acknowledgements

The LUCAS topsoil dataset used in this work was made available by the European Commission through the European Soil Data Centre managed by the Joint Research Centre (JRC), http://esdac.jrc.ec.europa.eu/. The author thanks the teams that measured and released the fractionation and spectral data and the predicted fraction map.

References

1. Begill N, Don A, Poeplau C (2023) No detectable upper limit of mineral-associated organic carbon in temperate agricultural soils. Glob Change Biol 29(16):4662–4669. https://doi.org/10.1111/gcb.16804
2. Breure TS, De Rosa D, Panagos P, Cotrufo MF, Jones A, Lugato E (2025) Revisiting the soil carbon saturation concept to inform a risk index in European agricultural soils. Nat Commun 16(1):2538. https://doi.org/10.1038/s41467-025-57355-y
3. Breure TS, Jones A, Panagos P (2026) Evaluating visible near-infrared spectroscopy in context of a repeated sampling survey across the European Union. Geoderma 465:117647. https://doi.org/10.1016/j.geoderma.2025.117647
4. Brus DJ, Kempen B, Heuvelink GBM (2011) Sampling for validation of digital soil maps. Eur J Soil Sci 62(3):394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x
5. Champiny RE, Georgiou K, Lin Y (2026) Methods matter: examining the apparent saturation of soil mineral-associated organic carbon. Geoderma 467:117732. https://doi.org/10.1016/j.geoderma.2026.117732
6. Cotrufo MF, Ranalli MG, Haddix ML, Six J, Lugato E (2019) Soil carbon storage informed by particulate and mineral-associated organic matter. Nat Geosci 12(12):989–994. https://doi.org/10.1038/s41561-019-0484-6
7. Field CA, Welsh AH (2007) Bootstrapping clustered data. J Royal Stat Society: Ser B (Statistical Methodology) 69(3):369–390. https://doi.org/10.1111/j.1467-9868.2007.00593.x
8. Georgiou K, Jackson RB, Vindušková O, Abramoff RZ, Ahlström A, Feng W, Torn MS (2022) Global stocks and capacity of mineral-associated soil organic carbon. Nat Commun 13(1):3797. https://doi.org/10.1038/s41467-022-31540-9

9. Hassink J (1997) The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil 191(1):77–87. https://doi.org/10.1023/A:1004213929699
10. Lavallee JM, Soong JL, Cotrufo MF (2020) Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. Glob Change Biol 26(1):261–273. https://doi.org/10.1111/gcb.14859
11. Lugato E, Lavallee JM, Haddix ML, Panagos P, Cotrufo MF (2021) Different climate sensitivity of particulate and mineral-associated soil organic matter. Nat Geosci 14(5):295–300. https://doi.org/10.1038/s41561-021-00744-x
12. Panagos P, Van Liedekerke M, Borrelli P, Köninger J, Ballabio C, Orgiazzi A, Montanarella L (2022) Soil data and knowledge in support of the EU policies. Eur J Soil Sci 73(6):e13315. https://doi.org/10.1111/ejss.13315. European Soil Data Centre 2.0:
13. Piñeiro G, Perelman S, Guerschman JP, Paruelo JM (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling 216(3–4):316–322. https://doi.org/10.1016/j.ecolmodel.2008.05.006
14. Poeplau C, Don A, Six J, Kaiser M, Benbi D, Chenu C, Nieder R (2018) Isolating organic carbon fractions with varying turnover rates in temperate agricultural soils – A comprehensive method comparison. Soil Biol Biochem 125:10–26. https://doi.org/10.1016/j.soilbio.2018.06.025
15. Six J, Conant RT, Paul EA, Paustian K (2002) Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant Soil 241(2):155–176. https://doi.org/10.1023/A:1016125726789
16. Stenberg B, Rossel V, Mouazen RA, A. M., Wetterlind J (2010) Visible and near infrared spectroscopy in soil science. Adv Agron 107:163–215. https://doi.org/10.1016/S0065-2113(10)07005-7
17. Stevens A, Nocita M, Tóth G, Montanarella L, van Wesemael B (2013) Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. PLoS ONE 8(6):e66409. https://doi.org/10.1371/journal.pone.0066409
18. Tóth G, Jones A,Montanarella, L. (eds) (2013) *LUCAS Topsoil Survey: methodology, data and results*. JRC Reference Report, EUR 26102 EN. Publications Office of the European Union, Luxembourg. https://doi.org/10.2788/97922

Figures

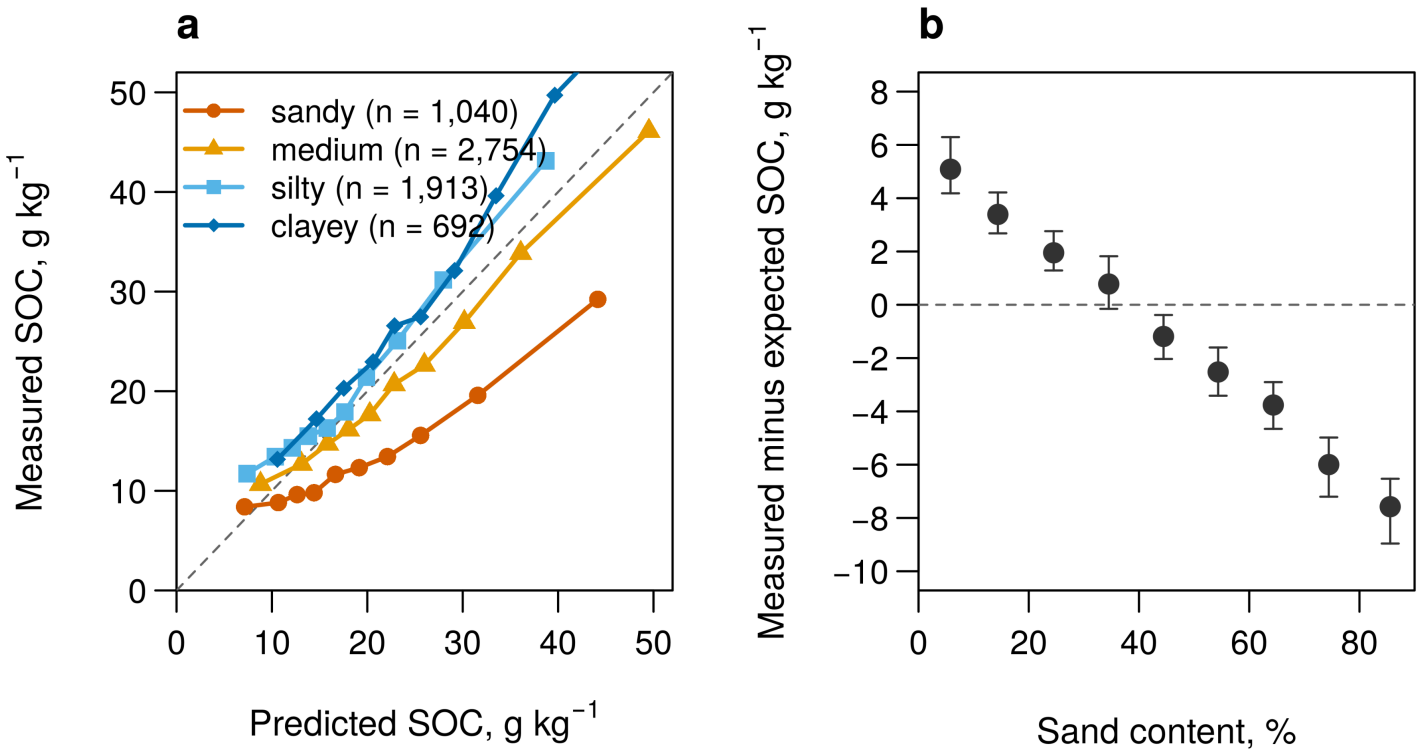


Figure 1

Measured soil organic carbon (SOC) against the summed predicted particulate and mineral-associated organic carbon of a published vis–NIR map at 6,399 LUCAS 2009 points. (a) Mean measured against mean predicted SOC in deciles of the predicted value within each texture class. (b) Mean departure of measured SOC from the regression of measured on predicted SOC over all points, by class of sand content, with 95% intervals from 1,000 bootstrap resamples of 2° grid cells.

Note: the dashed line in (a) is the 1:1 line; points below it are over-predicted. Texture classes: clayey, clay ≥ 35%; sandy, sand ≥ 65%; silty, silt ≥ 50%; medium, the remainder. A positive value in (b) means that measured SOC is higher than the overall regression expects from the prediction. vis–NIR, visible–near-infrared.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryCode.zip