

Point-by-point Responses to Editor and Reviewers

Manuscript egusphere-2025-3440: “Soil Health Management Outweighs Edaphic Properties in Explaining Soil Organic Matter Across Working Organic Farms”

We thank the editor and reviewers for their thorough and constructive feedback on our manuscript. Their comments substantially improved the rigor, clarity, and transparency of this work — most notably prompting us to identify and correct a back-transformation error affecting all model-based effect sizes in the original Table 2, and to substantially revise our framing of statistical significance and causal language throughout. Below, we respond to each comment in turn. Reviewer comments are shown in italics; our responses follow, with the relevant manuscript section noted in brackets.

Response to the Editor

***Comment:** Currently, your claims remain stronger than your evidence. For instance, this study appears to be observational; however, you use causal language in several places. Please consider moderating your claims.*

Response: We conducted a full pass through the Title, Abstract, Results, Discussion, and Conclusion, replacing causal verbs (“drives,” “increases,” “significantly increases,” “boosts”) with associational language (“is associated with,” “corresponds to”). The title was changed from “... Drives Soil Organic Matter More Than Edaphic Properties...” to “... Outweighs Edaphic Properties in Explaining Soil Organic Matter...”. We also added an explicit statement in §4.1 that “direct causal relationships and mechanisms between management and carbon outcomes are challenging to parse from this type of on-farm work; the associations reported here should be read as consistent with, rather than proof of, the mechanisms discussed.” [Title; Abstract; §4.1, §4.2, §4.3, §4.5 (throughout)]

***Comment:** The main problem is that the study compares only 28 fields at one point in time. It can show that certain farming practices and higher soil carbon occur together, but it cannot prove that those practices caused the higher carbon.*

Response: We agree and have revised the manuscript throughout to describe associations rather than causal effects. We also added an explicit statement of this limitation to §4.1: our sample size and cross-sectional design limit our ability to test interaction effects and to establish causal mechanisms. [§4.1]

***Comment:** The authors also claim to test whether soil properties limit the effects of management, but they do not actually test whether the same practice works differently across soils.*

Response: We revised one of our research questions in the Abstract, and subsequently in Introduction, from “How much do inherent soil properties limit the effectiveness of recommended soil health practices in increasing SOC?” (which implies an interaction/moderation test we did not run) to what was previously in the introduction: “How do soil health practices compare to edaphic factors in influencing carbon fractions and overall stocks?”, which accurately describes our variance-

partitioning approach. We did not add interaction terms, given our sample size (n=28 sites); we state this explicitly as a direction for future work in §4.1. [Abstract; §4.1]

Comment: *Several measurements and calculations must be checked. For example, the equations for particulate carbon do not clearly align with how the results are described; the concentration of mineral-associated carbon is treated as though it were the total amount of carbon stored; and bulk density is used in a way that may partly build the expected result into the analysis.*

Response: We addressed each of these three points directly. First, we added Table 2a/2b, providing a complete, explicit reference table of every equation and variable used to calculate fraction proportions, concentrations, and stocks, resolving the ambiguity between %C, C_{prop}, and stock values. Second, we added an explicit statement in §2.3 that MAOM is modeled as carbon concentration (MAOM C%) rather than total MAOM carbon stock, and committed to using “MAOM C%” consistently throughout the Results and Discussion to avoid this conflation. Third, we directly tested whether bulk density differs systematically across our management gradients (a prerequisite for the fixed-depth stock comparison to be unbiased); we found no consistent directional trend in surface bulk density across terciles of either the continuous-cover or disturbance gradient (Table S10), and report this check explicitly in §4.2. [§2.3; Table 2a/2b; §4.2; Table S10]

Comment: *Some management measures are too weak for the claims being made: satellite images cannot reliably distinguish cover crops from cash crops or weeds, and the use of amendments in a single season cannot offset soil carbon accumulated over many years.*

Response: For the cover-crop measure: we clarified in §2.3.2 that our primary continuous-cover variable intentionally includes both cash and cover crops, consistent with the “continuous living roots” mechanism we hypothesize (living roots from any crop are expected to contribute the relevant belowground inputs). We separately acknowledge, for our narrower winter-cover-specific metric, that satellite imagery could not distinguish dedicated winter cover cropping from an extended cash-crop season. For amendments: we elevated this from a post-hoc explanation to an a priori caveat, now stated in §2.3.3 before results are presented: organic amendment use was assessed for the 2020 season only, and even a representative annual rate may not capture the cumulative, multi-year amendment history most relevant to detecting an SOC response. [§2.3.2; §2.3.3]

Comment: *In addition, Table 1 contains impossible values, some confidence intervals in Table 2 appear implausibly narrow, and parts of the analysis seem to have been chosen after the authors saw which results were most favourable.*

Response: Table 1: the reversed min/max values (Bed Permanence, Deep Tillage Frequency, Tillage Depth) have been corrected against the raw survey data. Table 2 (now Table 3): we identified and corrected a back-transformation error in our original code, which applied the inverse link function (log, logit, Box-Cox) to model coefficients in isolation rather than to the full linear predictor before differencing — valid only for an identity link. This produced artificially narrow confidence intervals. We rewrote the back-transformation using simulation from the fixed-effects covariance matrix (described in §2.4), and all values in Table 3 reflect this correction; several results that originally appeared strongly significant (e.g., MAOM associations with continuous cover and tillage) are now reported as weaker, borderline associations (p = 0.07–0.11) rather than being overstated. Model

selection: we added an explicit statement of our pre-specified decision rule for when an individual-practice model is adopted over a composite-score model (based on random forest importance, directional consistency with established mechanisms, and model fit), rather than leaving this implicit. [§2.3.3 (Table 1 source); §2.4, Table 3; §2.3 (model-selection rule)]

Comment: *The authors reported sampling only the upper 15 cm of soil, while tillage depths exceeded 1 m. Tillage may simply move carbon deeper in the soil rather than destroy it. The authors must either provide evidence from deeper soil or restrict every conclusion to the surface soil. Carbon-stock comparisons should also account for differences in soil density rather than compare only a fixed depth.*

Response: We added a new depth-resolved analysis (§2.4.1, §3.6) using bulk carbon measured at three depth increments (0–15, 15–30, 30–60 cm) at harvest. We found no evidence of a surface carbon deficit at more-disturbed sites, which would be expected under a simple redistribution mechanism; if anything, less-disturbed sites showed higher carbon at 30–60 cm, a pattern confirmed across the full disturbance gradient (not just a single site-pair comparison) and not attributable to more favorable edaphic conditions at those sites (Tables S7–S9). We report this as suggestive rather than conclusive, since it uses bulk rather than fraction-specific carbon. We also explicitly restrict fraction-level tillage conclusions to surface soil in our Discussion and Limitations text (§4.1). For the density concern: as noted above, we tested and found no systematic difference in measured surface bulk density across our management gradients (Table S10; §4.2), addressing the specific concern that a fixed-depth comparison might conflate carbon concentration with differences in sampled soil mass. [§2.4.1, §3.6, Tables S7–S9; §4.1 (surface-soil scope statement); §4.2, Table S10]

Comment: *The revised paper must correct and clearly document all calculations, provide the full data and computer code, simplify and justify the statistical analysis, treat the field, not each soil sample, as the main independent unit, and rewrite the title, abstract, discussion, and policy claims so that they describe uncertain relationships found in these 28 organic lettuce fields rather than proven causes or broad climate benefits.*

Response: Calculations: addressed above (Table 2a/2b; back-transformation correction). Data and code: we added a Data and Code Availability statement; the underlying data and code are currently embargoed pending publication of companion manuscripts sharing this dataset, consistent with SOIL's data policy for embargoed datasets, and are available to reviewers and editors on request in the interim. Statistical analysis: we added an explicit, pre-specified model-selection rule (above) and clarified our imputation method, which we found was inadequately described in the original submission as mean substitution. Field as unit of replication: we added an explicit statement of sample size (n=138 observations across 28 sites) and intraclass correlation coefficients for all six primary models (ICC = 0.33–0.87), demonstrating that a substantial share of variance lies between rather than within sites, supporting the field as the appropriate unit of inference.

Title/Abstract/Discussion/Policy: revised throughout as described above. [§5 (Data and code availability); §2.3 (imputation, model selection); §2.3.2 (ICC, sample size)]

Response to Reviewer 1

Comment: *I would suggest putting together a small table like that used in stable isotope studies... It would just help make the paper much more accessible.*

Response: We added Table 2a (equations) and Table 2b (variable definitions), providing a complete reference for how every carbon pool, proportion, and stock is calculated. [Table 2a, Table 2b]

Comment: *I'm assuming you looked at surface soils and I think it would be good to just emphasize or connect directly that this reflects your lettuce root depths... it's not clear to me if this also encompasses the depth of your cover crop plants... you don't explain what you mean by functional types of crops, define this.*

Response: We added a direct statement connecting our 0–15 cm sampling depth to lettuce's shallow rooting depth, and explicitly acknowledge that cover crop rooting depths may extend below this. We also added a definition of cover crop functional groups, based on the USDA-NRCS's six defined groups (cool/warm season grasses, legumes, and broadleaves), with a citation. [§2.1.2; §2.3.1]

Comment: *Line 60: You mention MAOM being important for nutrient provisioning but don't support this with literature... Similarly, you describe POM fraction C/N ratios but don't reflect on MAOM ratios.*

Response: We added the two suggested citations (Liptzin & Silver, 2009; Spohn, 2024) supporting the MAOM nutrient-provisioning claim, and added a sentence on MAOM's typically lower C:N ratio relative to POM, citing Jilling et al. (2018) as suggested. [§1 (Introduction)]

Comment: *Is the mass of the fPOM and oPOM represented in these equations for the mineral+carbon+soil content combined or is this just C mass?*

Response: This is now unambiguous in Table 2b, which explicitly defines mfPOM and moPOM as the total (whole) mass of each fraction, distinct from the carbon mass within it (mC,MAOM), and clarifies that mtotal refers to the total bulk soil sample mass used for fractionation. [Table 2b]

Comment: *Line 170: You air dried your soils to determine the size of various Fe pools but air drying can lead to irreversible Fe mineralogical changes... no edit necessary, I just wanted to make sure this was also something that the writers are aware of.*

Response: We appreciate this note and have taken no action, per the reviewer's own indication that no edit was necessary; we are aware of this consideration for future sampling designs. [No change (reviewer indicated no edit necessary)]

Comment: *On Table 1 for Organic Amendments (Total Carbon Input) There seems to be an error, because the Minimum value (4489) is higher than the Max value (2991) – same for the Reduced Disturbance categories, all min values are higher than the max value listed.*

Response: We corrected the reversed min/max values for Bed Permanence, Deep Tillage Frequency, and Tillage Depth against the raw survey data. All rows in Table 1 now show $\text{Min} \leq \text{Max}$. [Table 1]

Comment: *Apologies if this was missed in my review, but I don't believe your time variable was integrated into your analysis? You mention sampling three times in the year... it isn't clear in your current manuscript what you did with this resolution of data you collected.*

Response: We clarified in §2.1.2 that soil health management effects on carbon fractions were assessed using mid-season samples only, and that bulk (unfractionated) carbon was additionally measured at harvest across three depth increments, used in a separate depth-resolved analysis (§2.4.1, §3.6). The transplant timepoint was not used in the carbon fraction analyses; we now state this explicitly rather than leaving it implicit. [§2.1.2]

Comment: *You refer to looking at %C in MAOM and noting that this is not the same as the total MAOM C stock. However, it's not clear from your description or written equations what the difference is... It might behoove the writers to add a table in the supplement where % C in X pool, absolute C in X pool, pool Y stocks are all defined.*

Response: Addressed via Table 2a/2b (main text, per our judgment that this table's importance warranted inclusion in the main text rather than supplement) and an explicit statement in §2.3 that we use “MAOM C%” consistently to refer to concentration, distinct from stock, throughout the Results and Discussion. [§2.3; Table 2a/2b]

Comment: *I don't think it's clear from the description that PC1 and MAOM2 were included to be an edaphic variable grouping in Figure 3 description. Please rewrite it to be clearer.*

Response: The Figure 3 caption (Marginal R^2 values) has been rewritten to explicitly state that the edaphic/biological grouping includes soil physical PC1, soil iron PC1, and pH for all models, and additionally includes soil enzyme activity PC1 for the MAOM2 model. [Figure 3 caption]

Comment: *Line 365: You introduce the term individual practice variables... but you don't ever define what individual practice variables are... Please define these earlier on in the text.*

Response: We now define “individual practice variables” at first use, distinguishing them from the composite management z-scores, and clarify (where relevant) that the specific variable substituted differs by carbon pool (e.g., tillage depth for oPOM vs. deep tillage frequency for fPOM). [§2.3, §3.3, §3.4]

Comment: *Line 395... you should include a line that points out at root redox processes that may be allowing iron mineralogy to transform and promote the sticking of aggregates to form oPOM.*

Response: We added a mechanistic paragraph to §4.2 on root-driven mineral transformation (root activity increasing availability of short-range-order Fe/Al phases that act as carbon-retention nucleation sites), citing Yu et al. (2017), and note this as a candidate pathway linking management-driven root inputs to iron-associated oPOM accumulation. [§4.2]

Comment: *Line 435, the sentence structure here makes this difficult to parse... Also there are some studies on accounting for interactions between soil health/intervention practices in grasslands, maybe include Shangshi paper on this?*

Response: The sentence has been restructured for clarity, and we added the suggested citation (Liu et al., 2025) as a precedent example of interaction-effects testing in a comparable multi-intervention context. [§4.1]

Comment: *Isn't Fe a micronutrient included in Soil Health Tests for metals? It's just that the kind of mineralogy that is tied to the presence of Fe and its ability to stabilize C isn't a measurement.*

Response: We revised the relevant sentence in §4.2 to state explicitly that total elemental iron is commonly measured as a micronutrient in routine soil health tests (Table S1), while the iron mineralogy relevant to carbon stabilization is not, due to the far greater labor intensity of sequential selective extraction relative to a standard elemental Fe test. [§4.2]

Comment: *Line 465, pyrophosphate capitalized when it shouldn't be*

Response: Corrected. [§4.2]

Comment: *Line 470, edaphic factors like texture (moisture/pore space) and pH may change dynamically over growing seasons... I'd be very curious to know if you get the same trends if you use soils collected from different periods.*

Response: We appreciate this observation as a direction for future work; as the reviewer did not request a specific textual change, we have not modified this section, though our new depth- and timepoint-scoping statements (§2.1.2) provide additional clarity on which timepoints and depths were used for which analyses. [No direct change requested; see §2.1.2 for related scoping clarifications]

Comment: *Line 490 what does directionality mean here?*

Response: We rewrote this passage to state explicitly that the ambiguity concerns the direction of causality between enzyme activity and carbon pools (e.g., whether enzymes drive substrate breakdown or respond to substrate availability), rather than the sign of the correlation, and explain why we treat this differently for POM versus MAOM. [§4.2]

Comment: *What mechanisms generate formation of these different pools, why is it that oPOM and MAOM seem to have formed with strong influence from total Fe variables... Yu, G., Xiao, J., Hu, S... (2017).*

Response: Addressed via the new root-driven mineral transformation paragraph described above, citing Yu et al. (2017). [§4.2]

Comment: *Root exudates and enzymes contribute to MAOM formation however Hui Li's paper on simple metabolites does not show this, this paper is focused on destabilization/disruption, which I don't think this is an appropriate use of the citation here. I would refer to some of Sokol's papers for references on MAOM formation from root products. Also: Liang, G., Stark, J., Waring, B.G., 2023... Other references... Klink, S... (2022)... Wang, T., Tian, Z... (2017).*

Response: We revised the relevant passage in §4.2 to cite Sokol et al. (2019) and Liang et al. (2023) specifically for the MAOM-formation claim, and Klink et al. (2022) and Wang et al. (2017) for fungal contributions to MAOM formation, as suggested. H. Li et al. (2021) and Jilling et al. (2021) are now cited only for the destabilization mechanisms they actually demonstrate, not for formation. [§4.2]

Response to Reviewer 2

Reviewer 2 provided both a general summary (nine numbered points) and detailed, location-specific annotations on the manuscript PDF. We address the general points first, followed by the specific annotations grouped by manuscript section.

General points

Comment: 1. *Methodological Clarity: The methodology lacks sufficient detail...*

Response: We substantially expanded methodological detail throughout §2, including: an explicit equations/variable-definition table (Table 2a/2b); field collection technique for bulk density; the auger-based sub-sampling and compositing procedure; clarification of drying method (air-dried, not oven-dried, correcting an inaccuracy in the original submission); explicit statement of sample size (n=138, 28 sites) and intraclass correlation coefficients; a corrected, precise description of our imputation method (CART-based multiple imputation with pooling, not mean substitution); and an explicit, pre-specified model-selection rule. [§2 (throughout)]

Comment: 2. *Vague Language: There are several instances of vague language...*

Response: We revised numerous specific instances flagged in the annotated PDF (detailed below), and conducted an additional pass removing weak paragraph-opening transitions (e.g., “Thus,” “Additionally,”) flagged as a recurring pattern. [§1 (throughout)]

Comment: 3. *Abstract and Data Presentation: The abstract does not provide specific data or key results...*

Response: We rewrote the Abstract to include specific effect sizes, confidence intervals, and p-values for every named result (e.g., fPOM: +3.16 Mg C ha⁻¹, 95% CI 0.82–5.92), and to accurately distinguish results with strong versus borderline statistical support, consistent with our corrected Table 3 values. [Abstract]

Comment: 4. *Survey Data: ...the manuscript does not present any data generated from this survey.*

Response: Survey-derived management variables (cover cropping, tillage, crop diversity, and amendment use) are summarized with min/max/mean/median values in Table 1. We believe this addresses the substance of this comment. [Table 1; §2.3.1]

Comment: 5. *Site Selection: The selection of research sites is inadequately described...*

Response: We expanded §2.1.1 to describe farmer recruitment (USDA Organic Integrity Database, technical assistance providers), the rationale for restricting the sample to organic farms (to avoid confounding from synthetic fertilizer/pesticide effects on soil microbial activity and carbon dynamics), and clarified that field size varied but the sampling protocol was applied consistently regardless. [§2.1.1, §2.1.2]

Comment: 6. *Management Practices:* The manuscript fails to indicate the management practices employed...

Response: Management practices and their summary statistics are presented in Table 1, organized by the four soil health principle categories (continuous living cover, crop diversity, reduced disturbance, organic amendments). [Table 1; §2.3.3]

Comment: 7. *Data Collection Points:* Data were collected at three time points... while data were gathered from three soil depths, the manuscript does not show this depth-specific data...

Response: We added a full depth-resolved analysis using bulk carbon at three depth increments (§2.4.1, §3.6, Tables S7–S9), and clarified that fraction-level analyses used mid-season, surface (0–15 cm) samples only, explaining this scope explicitly rather than leaving it unstated. [§2.1.2, §2.4.1, §3.6]

Comment: 8. *Implications and Conclusions:* The implications and conclusions drawn from the study remain vague...

Response: We substantially rewrote §4.5 to open with a clear, specific statement of findings (which practices were associated with which pools, and with what statistical confidence), rather than only describing study design, and to close with a concrete, evidence-grounded statement of practical relevance rather than an open-ended call for more research. [§4.5]

Comment: 9. *Overall Quality:* ...the manuscript suffers from poor writing quality, with some instances of repetition...

Response: We conducted a full-manuscript proofreading pass along with numerous smaller typographical and grammatical corrections throughout.

Specific annotated comments (by section)

Comment: *Abstract:* “working lands” undefined; vague results language; “is this a pool?” on carbon stock.

Response: Replaced “working lands” with “agricultural lands” on first use; rewrote all result sentences with specific values (see general point 3, above); clarified that total carbon stock is reported separately from, not as one of, the named carbon fraction pools. [Abstract]

Comment: *Introduction:* ambiguity in “organic matter and improve soil health”; oPOM/POM sentence flagged “convoluted”; MAOM turnover claim disputed (“I don't think this is necessarily true... turnover of MAOM can be within less than a season”); enzyme/MAOM formation sentence unclear; edaphic/POM stability message unclear; “emphasizing” unclear; paragraphs beginning with “Thus,”/“Additionally,”; confounding-effects rationale unclear.

Response: Added “soil” before “organic matter” for clarity; restructured the POM/oPOM passage to define the general POM category before its fPOM/oPOM subdivision; revised the MAOM persistence claim to acknowledge a fast-cycling fraction, citing Jilling et al. (2025), a paper specifically demonstrating this point; restructured the MAOM-formation/enzyme sentence into two

sentences separating composition from mechanism; added an explicit topic sentence to the edaphic/POM-stability passage stating that the two described mechanisms act in opposite directions; replaced “emphasizing” with more concrete language; removed both flagged paragraph-initial transition words; and rewrote the organic-farm-selection rationale to state the actual confounding mechanism explicitly, with supporting citation. [§1 (Introduction, throughout)]

Comment: *Methods: field sizes not stated; sub-sample/transect description unclear; mid-season timing rationale unclear; drying method/duration unclear; fractionation method “not clearly described”; wet-sieving procedure unclear; recovery-verification method unclear; bulk density field collection method unclear; survey “relevant parts” vague; polyculture inclusion unclear; “SOM” undefined; “gradient of implemented soil health practices” not spelled out; “to some extent” and “wide area” flagged as vague.*

Response: We revised §2.1.1–§2.2.2 to address each of these: field size variation is now stated explicitly alongside farm-scale data; the sub-sample collection procedure now specifies auger use, spatial extent (within ~5 m, across several rows), and compositing method; mid-season timing is now described as identified in consultation with farm partners rather than via a standardized phenological measurement; drying method was corrected from an inaccurately stated oven-drying step to air-drying (confirmed with co-authors), resolving the ambiguity about duration; the fractionation description now leads with the Haddix et al. (2020) citation covering the wet-sieving and recovery-verification procedures; bulk density field collection now describes the ring being driven into and excavated from the soil; the organic-lettuce and geographic-sampling rationale sentences were rewritten to state their logic explicitly rather than relying on vague phrasing. [§2.1.1, §2.1.2, §2.2.1, §2.2.2, §2.3.1]

Comment: *Table 1: min/max value errors (Total Carbon Input, Reduced Disturbance categories).*

Response: Corrected against raw survey data; see response to Reviewer 1 on this same point. [Table 1]

Comment: *Results: units missing on CEC values; “what’s the message here” on the climate/edaphic variation paragraph; confusion over what the modeled MAOM percentage-change values (e.g., “~1.85%, CI: 1.79–1.90”) actually mean.*

Response: Added units (Meq/100g) to CEC values; rewrote the climate paragraph to state explicitly that shared regional climate across sites means climate is unlikely to confound the management/edaphic relationships under study. The confusion over modeled MAOM values will hopefully be aided by better description of the simulation-based method used to generate them, and are now described explicitly in §2.4 and reported with paired point estimates and 95% CIs throughout §3. [§3.1; §2.4, §3.2–§3.5, Table 3]

Comment: *Discussion: opening sentence flagged “meaning?”*

Response: This sentence, and the surrounding paragraph, were substantially rewritten as part of our broader associational-language revisions (see response to the Editor). [§4.1]

Comment: *“These results were not shown” (survey data); “There are no clear conclusions here” (Conclusion).*

Response: See responses to survey points above. Conclusion was substantially redrafted to center primary conclusions. [Table 1; §4.5]

We thank the editor and all three reviewers again for their careful engagement with this manuscript. We believe the revisions described above have substantially strengthened the rigor and transparency of the work, most importantly through the correction of the back-transformation error affecting all reported effect sizes, and we hope the revised manuscript now meets the standard for publication in SOIL.