Reviewer 1

The presented manuscript deals with the question why there is generally more organic C (SOC) in soils under aspen than under spruce trees in North America. The strength of this work is that its methods are quite comprehensive (and described with enough details), which makes it possible to assess complex interactions. A relative weakness of the study is the limited number of sites and especially the fact that aspen and spruce sites have (naturally) different positions in the landscape, which makes it impossible to extract the "pure" effect of the tree species. The authors are perfectly honest in acknowledging this fact and the discussion is written accordingly. The main conclusions are therefore not so much based on direct proofs but rather on clusters of indicators.

The article is generally well written (for a reviewer using English as a third language). There are, however, several long, sometimes complicated sentences that would deserve a simplification (see details below).

We greatly appreciate the feedback from the reviewer and have addressed the long sentences and simplified the text as indicated below.

The bibliography list is extremely long. Could it be shortened without affecting the understandability of the text?

We removed several references including but not limited to: Kelly et al., 2001; Larson et al., 2007; Jiang et al., 2023; Neubauer et al., 2013; Tange and Johannesson 2003; Pourret et al., 2007; Zeng et al., 2021

Details

L. 44: more SOC under aspen than under spruce: is this per soil mass or per area? Or both?

Changed to “we observed greater SOC concentrations under Aspen”, see Line 44.

L. 49: repeating "root" not necessary.
Corrected

L. 66: "SOC regulates" may be misunderstood as SOC being the sole factor regulating these properties.
Changed to influences

L. 102: the effect of coarse roots on aggregates is certainly small simply because there are by far less coarse than fine roots.
This may be true; indeed, as the reviewer states coarse root abundance was dwarfed by fine roots. However, our sentence is describing known impacts of coarse vs fine roots on aggregate collapse and formation, so we would like to keep it as is. Further, we cannot know in these soils whether few coarse roots impose less of an influence on aggregates than many fine roots.

L. 142: the word "tandem" is mostly used for a two-fold combination, i.e. its use here is not wrong but a bit surprising.
Changed to “at the same time”

L. 180: there are several shapes, therefore rather "shapes represent" (plural). The whole legend of this figure is quite long.
Changed to: “Shapes represent”. We appreciate that the caption is long, but due to the number of environmental factors that differ across the sites we were intentional in making the caption explicit so readers did not need to hunt through the document to understand how sites differed.

L. 193–194: it's not clear what "average yearly minimum and maximum" are. From the wording itself, it should be taking each year the minimal/maximal recorded temperature and then averaging over years. The given range is, however, very narrow for that, even more for a continental climate. Are these perhaps minimum and maximum monthly averages, then averaged over years?
We simplified this sentence to: “The mean annual temperature is 0.9 °C and the mean annual precipitation is 670 mm (Carroll et al., 2018), with approximately 60% falling as snow between October and May.”

L. 226: how many coring locations are there within these 100 m?
Sentence was changed to: “Twice in the summer of 2022 (late June and mid-August), soil was collected from 3 auger sampling locations within ~100 meters of each pit to characterize soil chemistry”.

L. 235: I would always encourage to refer to concentrations by using the word concentrations, and not indirectly by using the measurement unit like %N or %C.

One could actually also express the same parameter in e.g. g/kg. Note also that the C:N ratio is not "measured" but calculated.

We altered the text to address the comment, now L241: “We assessed SOC and SON concentrations and stocks and the likelihood of SOC degradation by microbes by analyzing bulk soil samples at 10-cm intervals. We determined SOC and SON on subsamples (~75 mg) via an elemental analyzer (Vario Macro Cube, Elementar, Ronkonkoma, NY). We used SOC and SON concentration measurements to calculate each subsample’s C:N ratio. To determine stocks of SOC in each horizon, we multiplied SOC concentrations by soil bulk density obtained in each horizon. Bulk density was measured using a three-dimensional laser scanner (3D Scanner Ultra HD, NextEngine, Inc., Santa Monica, CA) following Rossi et al. (2008).”

L. 242: it would be better to specify from the beginning on that this is extractable DOC (by opposition to a DOC concentration that would be measured by lysimetry in the field).

Now line 247, changed to: “ We measured extractable, dissolved organic C (DOC) to estimate organic C that can be relatively easily mobilized and transported out of the soil profiles; note that this differs from DOC measured in soil porewater using lysimeters, and instead represents a salt-extractable pool. “

L. 256: prefer the SI unit Hz (s^{-1}) to rpm.

“4800 RPM” change to “80 Hz (s^{-1})”

L. 258 ff.: long sentence.

Now line 265, changed to: “To assess the degree to which soil microbial communities were generating exo-enzymes that catalyze soil organic matter decay and can provide C- and N-rich compounds, we quantified potential activity rates of two such enzymes. We measured activity of β-glucosidase and N-acetyl-β-D-glucosaminidase, herein referred to as BGase and NAGase, which are linked to microbial C (BGase) and N and C (NAGase) acquisition (Sinsabaugh and Moorhead, 1994; Allison et al., 2011, Stone et al., 2014), using 4-methylumbelliferyl β-D-glucopyranoside (for BGase) 4-methylumbelliferyl N-acetyl- β-D-glucosaminide (for NAGase) fluorescent tags. These tags were added to slurries made from approximately 1 gram of soil and pH-adjusted 50 mM sodium acetate.”

L. 271 ff.: as nitrate in soils is subject to strong variations with time, please specify the vegetation and hydrology conditions at sampling time.

We added the month pits were sampled into Table 1, where the extractable nitrate concentrations are indicated to be sampled from all five pits (e.g., Aspen and Spruce).

L. 316: consider writing "oven-dry" with an hyphen.
Done

L. 318: "2-4.76": an en dash should be used here instead of an hyphen (even if one could argue that this is the job of the typesetter and not of the authors...)
Thank you, corrected.

L. 331: how does ImageJ recognize roots? This is not trivial at all.
This is a manual procedure, ImageJ does not recognize the roots. We have updated the text to clarify we updated line 336: "We used ImageJ (Schneider et al., 2012) to overlay each image with a 1x1 cm grid. We then manually checked each 1x1 cm grid cell for the presence of a fine root (diameter < 1 mm) or coarse root (diameter > 1mm) and noted these presence/absence scores for each grid cell.".

L. 346: use the full word "minutes" or the abbreviation "min", but not "mins".
Corrected.

L. 352 ff: this calibration procedure is not clear. Transforming ppm to % should be just dividing by 10000. This works in all cases: if these ratios are for moles, for volume or for partial pressure.
Thank you for catching this, the sensors provide a millivolt reading which are converted to a %. Now Line 352 reads: "We converted O_2 from millivolt reading to % by adding calibrated values to the millivolt value of O_2. Each calibrated value was specific to the sensor installed and determined using atmospheric concentrations prior to installation."

L. 393: be a bit more specific here than just writing "more variable". The text should essentially be understandable also without accessing the supplementary material.
Text on line 399 now reads : "Clay, silt, and sand content at the aspen sites (AS and EAG; Fig. S1) and one of the conifer sites (ESG) exhibited little variation with depth (average 33.1% clay and 18.8% sand), while the two other conifer sites had a greater sand and lower silt and clay content , particularly at depths greater than 25 cm (SS and SG; Fig. S1)."

L. 445: "similar range across depth" is not really clear to me as the ranges vary with depth according to the previous sentence, and the value given here is per area, i.e. apparently cumulated over all soil depths.

Text was updated to be more specific. Now starting on Line 452, “Across all sites, SOC concentrations ranged from 46.0-62.6 mg g^{-1} near the surface (5 cm deep) to 4.8 to 29.0 mg g^{-1} at depth (Fig. 4a). SOC concentrations were generally higher under aspen compared to spruce sites ($p < 0.0001$; Fig. 4a), but LME models also suggest that the best fit model included a significant interaction between vegetation and depth ($p < 0.001$), suggesting that SOC declines with depth for both vegetation types but to a greater extent under spruce compared to aspen. Stocks of SOC ranged between 0.01 and 1.31 kg m^{-2} (Fig. 4b) these values did not exhibit consistent declines with depth or clear differences across cover type.”

L. 451 ff: the text goes from the topic "DOC vs. SOC" to a second topic "additive vs. interaction" then back to the first topic and finally again to the second one. Consider rearranging this.

We have edited the text to be simpler, starting on Line 459: “In contrast to SOC concentrations, DOC was generally higher under the spruce stands compared to aspen. Similar to SOC, a model including a significant interaction between vegetation and depth was the best predictor of DOC values ($p < 0.001$), likely reflecting variable DOC values at different depths in both vegetation types (Fig. 4c). The DOC:SOC ratio also exhibited a significant interaction between vegetation type and depth (Fig. 4d; $p = 0.0007$). As with DOC, this significant interaction likely reflects variable ratio values for each cover type across depths.”

L. 469–471: complicated sentence.

This section was update, starting on line 471: “Total soil nitrogen ranged from 0.2 mg g^{-1} at depth to 4.63 mg g^{-1} near the surface. A model including an interaction between vegetation type and depth was the best fit ($p = 0.003$; Fig. 5a). Aspen values were greater than those in spruce-dominated soils at all depths; the significant interaction implies that the decline with depth was greater in spruce soils.

L. 476: this range of C:N values goes from extremely low to quite high. Any comment about such a low value?

We edited the text to provide context, Line 481. "The spruce sites showed greater variation with depth with a similar mean value of 19.3 in the top 20 cm but widely variable values at the deepest points, ranging from 4.6 to 28.7 (Fig. 5c). A model including a vegetation and depth interaction was a meaningfully better fit than all simpler models ($p = 0.0008$), suggesting that the visibly greater variation in C:N with depth in spruce soils was a significantly different depth trend from the fairly constant aspen values. The lowest value measured, at depth in one of the spruce forests, is suggestive of soil organic matter highly-processed by microbial communities (Ziegler et al. 2017)."

Fig. 9: the increases in CO_2 should be expected to relate to the decreases in O_2 by a relatively constant respiratory quotient. This is not the case here. Is the respiratory quotient really so different or are there other processes, or some measurement errors?

We appreciate that you brought this to our attention, and the point here is not to discuss the apparent respiration quotient (ARQ) but instead to generally give an understanding of the range in data. This presentation of the data also does not allow us to assess the time component of how CO2 and O2 interact but rather there ranges relative to each other. Our data are suggestive of silicate weathering and the dissolution of CO2 gas into soil water, which can lead to an AQR much lower than 1 (Hodges et al., 2019). Values with similar ranges were observed in soils in Shale Hills and Garner Run (Hodges et al., 2019).

Hodges, C., Kim, H., Brantley, S.L. and Kaye, J., 2019. Soil CO2 and O2 concentrations illuminate the relative importance of weathering and respiration to seasonal soil gas fluctuations. *Soil Science Society of America Journal*, *83*(4), pp.1167-1180.

L. 539: it does not make much sense for me to compare aspen with granite. Or do you mean spruce?

Good catch, changed to: "the spruce"

Fig. 10: the aggregate size classes seem not to add to 1, i.e. there is also a non-aggregated fraction. It would be at least as interesting to know the proportion of non-aggregated soil than to give the proportions of the size classes.

We respectfully disagree. It is unclear what we would learn from information about the non-aggregated portion of the soil plotted by depth. More importantly, however, the portion that falls below the bottom sieve is not completely unaggregated and we would not be able to separate the unaggregated from the aggregated portion of that material using the standard procedure for wet-stable aggregate-size distribution that we followed. The way we have represented this data is also fairly standard [see section 2.6.2.3 in the standard method described by Nimmo and Perkins (2002)].

L. 564: the word "further" is not really wrong as the verb "suggests" follows the verb "indicate" on L. 559. However, the sentence is so long that it comes quite in a surprising manner here.
We rephrased this section, changed to "By integrating knowledge from biology, pedology, hydrology, and soil chemistry we were better able to understand how multiple factors interact to drive observed SOC patterns in aspen and conifer montane forests. Our data indicate that differences in SOC protection give rise to often observed patterns of elevated SOC storage in soils under aspen compared to those in conifer stands (Woldeselassie et al., 2012, Laganiere et al., 2013, Boca et al., 2020, Román Dobarco et al., 2021). Furthermore, our study suggests that aspen-dominated soils may experience enhanced degrees of microbial transformation of SOC, with the products of those transformations exhibiting a greater tendency to reside in relatively small aggregates and thus protect carbon to a greater degree (Fig. 12)."

L. 608 ff.: long sentence. This is a bit counter-intuitive as larger aggregates would be expected to split into small aggregates and thus be at least as well able to protect DOC than such small aggregates. Does this deserve any comment?
Simplified the sentence and added additional discussion. Changed to: "Literature hints that the larger size aggregates (Fig. 10c) and greater propensity for C to form large aggregates (Fig. 11b) observed in the spruce-dominated soils at our sites may be due to a greater abundance of particulate organic matter (POM) in spruce compared to aspen forest soils (Cotrufo et al., 2015; Cotrufo et al., 2019); this may be the case if spruce litterfall is more difficult to decompose.".

L. 648: the word "confounded" would be good if this became a problem for a regression analysis. Here it is just about describing the situation and the word "related" would probably be better.
Rephrased to: "In our study aspen cover co-occurs where soil temperatures are warmer"

L. 670 ff.: long sentence.
Rephrased to: “Increases in stream water DOC concentration can harm global water quality by altering light and thermal regimes, nutrient cycling (e.g., Morris et al., 1995; Cory et al., 2015), the transport and bioavailability of heavy metals (e.g., Dupré et al., 1999; Trostle et al., 2016), and creating harmful disinfection byproducts (Leonard et al., 2022).”

L. 723: I don't understand the use of the word "import" in this context.
Rephrased to: “can be important”