**Reviewer: 1**

**Manuscript Number: bgb-2022-207**

## General Comments

The paper presented by Dobarco and colleagues represents an ambitious undertaking to map the fractional contribution of three different organic matter fractions (MAOM, POM, and PyOM), building on previous work from Grundy et al. (2015) and Viscarra Rossel et al., (2019). Using fractional carbon data predicted via mid- and near- infrared spectroscopy and quantile regression random forest model, they produce a useful gridded dataset of MAOC, POC, and PyOC to a depth of 30 cm. Generally, I found the paper to be well-written and insightful and have only few comments which I detail below.

| Principal Criteria | Excellent (1) | Good (2) | Fair (3) | Poor (4) |
|---|---|---|---|---|
| **Scientific significance:** Does the manuscript represent a substantial contribution to scientific progress within the scope of Biogeosciences (substantial new concepts, ideas, methods, or data)? | | X | | |
| **Scientific quality:** Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)? | X | | | |
| **Presentation quality:** Are the scientific results and conclusions presented in a clear, concise, and well-structured way (number and quality of figures/tables, appropriate use of English language)? | | X | | |

## Review Criteria

1. Does the paper address relevant scientific questions within the scope of BG?
   Yes.
2. Does the paper present novel concepts, ideas, tools, or data?
   Yes, especially the published gridded data.
3. Are substantial conclusions reached?
   Yes, the spatial distribution of POM, MAOM, and PyOM across Australia presents a substantial step forward towards understanding controls on SOM formation.
4. Are the scientific methods and assumptions valid and clearly outlined?
   Yes.
5. Are the results sufficient to support the interpretations and conclusions?
   Yes.
6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?
   Yes.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution?
   Yes.
8. Does the title clearly reflect the contents of the paper?
   Yes.
9. Does the abstract provide a concise and complete summary?
   Yes.
10. Is the overall presentation well-structured and clear?
    Yes.
11. Is the language fluent and precise?
    Generally, yes.
12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?
    Yes.
13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
    Yes.
14. Are the number and quality of references appropriate?
    Yes.
15. Is the amount and quality of supplementary material appropriate?
    Yes.

**Line Comments**

Lines 41 – 43: It's not clear whether SOM or SOC is being discussed in this sentence. I think the sentiments expressed are true in both instances, but it's worth rephrasing to improve clarity.

Lines 47 – 49: Is this true for PyOC as well? What is the primary mechanism of preservation for pyrogenic organic matter if not some level of biochemical recalcitrance?

Line 74: In the MAOM fraction, what is the mechanism of preservation of PyOM in the MAOM fraction? Is it occlusion within microaggregates that are smaller than 60/53 µm? Or can PyOM form organo-mineral association? I think it this section it would be worth discussing this fraction in a little more depth.

Lines 77 – 78: This is kind of what I mean in my comment above, it seems contradictory.

Lines 88 – 89: It would be great to have a citation here for either how fractions can inform management, or how they can be incorporated into policy.

Lines 139-142: Could you add a discussion either here or later in section 2.3 related to the pre-processing of the data from the different libraries? Were the data smoothed and corrected using Savitzky-Golay or the like, and how did that differ across the different libraries? If they are raw spectra, please specify that they were received in that form.

Line 219-220: Are there data associated with the C recovery of the fractions that might explain some of the poor matching between TOC and fraction sums? Soluble and dissolved carbon can be lost throughout the fractionation process, which may bias predictions.

Line 235: Can you clarify the difference between ilr1 and ilr2 in this line? Looking through the equations and text I think I can piece it together, but it would be helpful to the reader to make it explicit.

Line 280: How much of the dataset was void-filled? Can you provide a percentage for total data interpolation across covariates?

Line 387 and on: As my fellow reviewer noted, there are inconsistencies in the reporting of significant digits and errors in the results section of the manuscript. At the risk of being redundant, I recommend the authors carefully check the figures they present for consistency and utility.

Line 498-499: Are most agricultural lands in the Mediterranean biome irrigated? It could be worth highlighting the proportion irrigated either here or in the discussion.

Line 517: Higher mean sand content or higher mean SOC concentration?

Line 522-523: I think this sentence is confusing, I recommend rewording. Currently it reads as if the authors are discussing the total proportion of MAOC across Australia.

Line 541: Please clarify the directionality of the relationship between POC and MAT.

Lines 552 - 554: I'm glad you mention this here -- I was going to recommend something along these lines as a justification for not including these co-variates.