Review of Houston et al. "Old Carbon, New Insights: Thermal Reactivity and Bioavailability of Saltmarsh Soils" (Biogeosciences; https://doi.org/10.5194/egusphere-2024-3281)

Synopsis

The primary focus of this study is to determine Ramped Oxidation (RO) ^{14}C activities (reported as "percent modern" or pMC) and $\delta^{13}C$ values for organic carbon (OC) from a set of saltmarsh soil cores from the Skinflats saltmarsh in Scotland, UK. The authors find that, in general, pMC decreases and $\delta^{13}C$ increases with increasing RO temperature in all studied samples. They interpret this result as evidence that saltmarshes store significant amounts of pre-aged, (thermally) recalcitrant OC (although I think the terms "labile" and "recalcitrant" get conflated with "bioavailable" throughout this study; detail below). The authors then compare these results with those of incubation experiments from the same set of cores (from an earlier study; Houston et al. 2024 *Limnol. Oceanogr.*), and conclude that, for most samples, respired OC matches most closely in its ^{14}C activity with that from the lowest thermal fraction. This is used as evidence that remineralizing organisms largely use thermally labile OC as substrate.

This is a concise manuscript, and I think the comparison between RO and biological incubation ^{14}C analysis is an interesting and underexplored area of research. I therefore find the overall theme of this manuscript interesting and fitting for *Biogeosciences*. That said, there are major weaknesses with the current manuscript that need to be addressed and fixed. I broadly define these as:

(i) a lack of citation and acknowledgement of the primary RO literature, which is largely overlooked here;
(ii) relatedly, poor framing of results within the existing RO data analysis and interpretation pipelines (i.e., no determination of activation energy, *E*, distributions);
(iii) a lack of detail on measurement and sample analysis (most importantly, that samples were not decarbonated prior to RO);
(iv) a lack of detail on methods validation and verification (e.g., measured bulk vs. RO mass-balance OC contents, ^{14}C activities, and $\delta^{13}C$ values);
(v) conflation of concepts and overall "sloppy" use of terminology (particularly labile / recalcitrant vs. bioavailable, as well reporting of ^{14}C ages for complex OC mixtures);
(vi) improper use of regressions and data analysis (i.e., exponential and linear regressions when the *x* axis is not properly reported as a continuous function).

I detail each of these issues below. I refrain from making line-item comments, as a large portion of this text will likely need to be be re-written. I believe solving these issues---particularly issue (ii), in which the authors should utilize the well-established framework for interpreting RO data that has been developed over the past ~decade---will greatly improve the strengh of conclusions that can be drawn here. Only after these issues are addressed can I assess and comment on the discussions, interpretations, and conclusions

Thus, I do not support publication of the current version of this manuscript in *Biogeosciences*. However, I outline below my recommendations for how the authors could revise and re-write to better focus on more interesting discussion and interpretation within the context of the broader knowledge in this field. I would then be happy to re-review a revised version. Please do not hesitate to contact me regarding any questions on this review.

Sincerely,

Jordon Hemingway
jordon.hemingway@erdw.ethz.ch

Major comments

1. Primary R(P)O literature and context

There exists a rich literature of ramped (pyrolysis) oxidation studies dating back to Rosenheim et al. (2008) *Geochem. Geophys. Geosys.* that is largely ignored here. While I understand the primary focus of this manuscript is on saltmarsh soils, it is important to place this within the broader context thermal analyses, particularly when introducing the RO instrument (e.g., L76-85, L109-125). This body of literature consists, for example, of studies related to:

(i) instrument design and underlying theory (e.g., Rosenheim et al. 2008 *Geochem. Geophys. Geosys.*; should be cited on L78 rather than Garnett et al. 2023);
(ii) blank assessment (e.g., Hemingway et al. 2017 *Radiocarbon*, which is cited here but in a different context; Fernandez et al. 2014 *Anal. Chem.*, etc.);
(iii) interpretive framework with respect to thermal activation energies and various OC sources (e.g., Hemingway et al. 2017 *Biogeosciences*, Hemingway et al. 2018 *Science*);
(iv) data compilations, particularly as they relate to mechanisms of OC preservation such as organo-mineral interactions (e.g., Hemingway et al. 2019 *Nature*, Cui et al. 2022 *Science Advances*.

There is admittedly some self-citation in this list, but my overall point here is that many of the studies that first described and developed the framework for such thermal oxdation instrumentation is ignored in the present study.

2. RO data analysis and interpretation

Perhaps more importantly, the current manuscript takes an overly simplistic approach to data analysis and interpretation---particularly, the authors simply "bin" data into temperature windows (i.e., 150-325 °C, 325-425 °C, 425-500 °C, 500-650 °C, and 650-800 °C) and further bin these into "labile" (i.e., 150-425 °C) and "recalictrant" (i.e., 425-650 °C) fractions to perform all analyses and interpretations. This is not robust. This becomes apparently when the authors attempt to perform regressions using these "bins" on the *x* axis despite the fact that they are not evenly distributed along temperature (see point 6, below).

Furthermore, the authors normalize all thermograms to maximum peak size within a given thermogram (i.e., such that the *y* axis scales from 0 to 1, inclusive) and perform regression analyses on these normalized bins. This is again not robust since, for example, one sample could contain a tall-but-narrow peak that may not be a large contributor in terms of overall area (and thus fraction of total OC content). Rather, to be properly compared, all thermograms should be normalized such that the integral under each is equal to unity.

Forunately for the authors, there exists a well-established data anlaysis pipeline for the quantiative interpretation and comparison of RO data, as described in Hemingway et al. (2017) *Biogeoscineces* and easily implemented using the "rampedpyrox" Python package. In this framework, thermograms are converted to activation energy, E, distributions, $p(E)$, and RO ^{14}C activities and $\delta^{13}C$ values are plotted vs. the weighted-mean E value for a given fraction. This approach allows for direct comparison between samples and datasets, and it has the added benefit of providing a continuous *x* axis for interpreting isotopic trends with increasing thermal recalcitrance (e.g., as is attempted in the current manuscript's Figs. 2-3). Additionally, placing these data within an activation energy framework significantly simplifies and strengthens many discussion points, for example those made throughout Section 4.2. For example, one can determine thermogram bredth using $p(E)$ width, σ_E, which allows for easy comparison of the

importance of thermal recalcitrance between samples (as is currently done in a somewhat "clunky" manner beginning on L231).

I therefore strongly recommend that the authors interpret their data in an activation energy context, rather than as temperature "bins" as is currently done. Doing so will significantly improve data interpretation and will lead to more quantitative and robust trends.

3. Analysis and measurement detail

Significantly more detail on the analytical setup and sample preparation is necessary. For example, the reader should be able to easily determine: sample masses, carrier gas composition and flow rate (the authors state "stream of high purity oxygen" on L113, but I seriously doubt it is pure O_2), CO_2 masses needed for each ^{14}C and ^{13}C measurement, calculated isntrument blank contribution, etc. All of these details need to be listed and described in order for the reader to be able to trust and interpret any results.

Additionally, the reader has to wait until Section 4.3 (L296) before learning that samples have not be decarbonated prior to RO analysis! This is a major oversight, as this is critically important information for the reader to have when interpreting data. Relatedly, how do the authors account for the potential of Inorganic Carbon (IC) to begin to combust at ~550 °C, as has been shown in e.g., Hemingway et al. (2017) *Radiocarbon*? Such a phenomenon would lead to an OC/IC admixture at higher temperature fractions, which could be an alternative mechanism to explain increasing $\delta^{13}C$ value with increasing temperature (c.f., microbial decomposition, which is proposed as the mechanism on L284). Overall, the authors need to be much more clear about their sample handling procedures (particulalry lack of decarbonation), and they need to thoroughly and honestly discuss how these procedures may impact their results and interpretation. Currently, I see none of this in the manuscript.

Finally, a small point, but the authors state that IRMS-derived $\delta^{13}C$ data were, "used to normalise the ^{14}C results to a $\delta^{13}C$ of -25 ‰ to correct for isotopic fractionation" (L123). Is this really true? If so, that is a major break from typical procedure, which is to use the AMS-derived $^{13}C/^{12}C$ ratio for correction, as this includes any internal fractionation (e.g., during ionization).

4. Data validation and verification

Similar to the above point, I find several data validation and verification metrics missing. For example, what are the bulk sample %OC, %OC, pMC, and $\delta^{13}C$ values? How well do the authors' RO results reconstruct these bulk values? That is, if the authors take a weighted-average of their RO fractions, are they able to reproduce the bulk values within statistical uncertainty? These types of "sanity check" metrics are quite important and, without them, I find it very difficult to assess the robustness and validity of the reported RO data. Again, fortunately for the authors, there exists a well-established and -documented pipeline for performing these types of sanity checks, and it is easily implemented using the "rampedpyrox" python package.

Finally, another small(ish) point, but the authors suggest ^{13}C fractionation as a possible reason for increasing $\delta^{13}C$ values with increasing combustion temperatures (L289-291). We actually show in the cited paper that fractionation is *not* an important factor and likely only shifts results by $\leq$ 1 ‰ (Hemingway et al. 2017 *Radiocarbon*). Rather, the authors could consider mechanisms such as different macro-molecular compounds (e.g., lipids vs. carbohydrates) combusting at different temperature ranges or the importance of organo-mineral interactions for some compound classes. Furthermore, as mentioned above, the possibility of carbonate

contribution as low as ~550 °C needs to be addressed here, as this would result in a similar $\delta^{13}C$ trend for the medium- to high-temperature fractions.

5. Concepts and terminology

There are several instances throughout the manuscript where I find the terms "labile" and "recalcitrant" to be conflated with "bioavailable" (e.g., L62: "It is therefore assumed that old OC is mostly composed of recalcitrant (low reactivity) components, whereas young OC contains a greater proportion of labile (reactive) components; L78: "The energy required to thermally-evolve CO_2 is expected to be related to the energy required for biological degradation of OC, with CO_2 evolved at low temperatures deemed to be from more reactive soil OC pools than CO_2 evolved at higher temperatures"; L203: "implying that the reactivity of soil OC decreases with increasing temperature", L205: "low-temperature CO_2 peak as relatively 'labile' and the higher temperature CO_2 peak as 'recalcitrant' OC pools", etc.).

It is clear that the authors know the difference between these concepts, but the language of the current manuscript is sloppy in a way that that could easily lead to reader conflusion. I strongly suggest the authors only refer explicitly to *thermal* lability and *thermal* recalcitrance---as this is what their RO instrument is directly measuring. Then, any relationship to OC *bioavailability* or *turnover time* can be inferred or interpreted, particularly using ^{14}C activities. However, it is important to clearly articulate that increased *thermal recalictrance* need not correlate with biological turnover times---the former is merely an analytical tool to parse apart complex OC mixtures, while the latter is the true metric of interest in the environment.

Similarly, at some points in the manuscript the authors report their ^{14}C activities as traditional ^{14}C years before present (L265, L276) and interpret them within the context of paleoclimatic events (e.g., deglaciation). This is very dangerous. The ^{14}C age in years BP of *any complex OC mixture* is meaningless, as this is simply the weighted average of all compounds contained within this mixture. Thus, any correspondence between, say, the ^{14}C age of a given RO thermal fraction and the deglaciation is pure coincidence---it does not mean that all (or even any) OC compounds that are represented in said RO fraction were formed at that time. I strongly suggest the authors remove this interpretation.

6. Regressions, plotting, and statistics

I have several comments and suggestions related to the figures and interpretations thereof:

Fig. 1: I suggest using something besides color to distinguish ^{14}C activities---previous studies have included overlaid bar plots or scatter plots, which convey this information much more clearly. For example, when I printed the manuscript in black and white, I could not distinguish the ^{14}C colors at all. Also, looking at the thermogram, it is clearly evident that carbonate is present in the T1 0.5cm sample, but this is only mentioned much later in the manuscript (Section 4.2) and not at all addressed in the figure itself. Finally, strictly speaking, the gray region does not "indicate values outside of the CO_2 collection range (150-800 °C)", as the gray region also includes the range 650-800 °C for all but one sample.

Fig. 2 + Fig. 3: Here, the authors are plotting RO isotope results vs. an *x* axis that is equally spaced temperature fraction means despite the fact that said temperature fractions are *not* equally spaced. Thus, any regression functional forms have no meaning (i.e., these are not, in fact, exponential and linear, respectively, if the *x* axis were to show temperature in a proper, continuous form). This needs to be fixed. Again, fortunately for the authors, there exists an

entire interpretative framework based on thermal activation energy distributions, $p(E)$, as well as a python package that makes this possible with very few lines of code.

Additionally, by plotting all tempererature fractions as box-and-whisker plots, the reader loses significant information *for a given sample*. That is, there is no way to know, for example, which point in the 150-800 °C bin corresponds to the same sample in the 325-425 °C bin. By doing this, the authors are inherently reverting to the mean values across all samples for a given temperature bin, which loses nuance and information (and may be incorrect depending on how OC is distributed *whithin* each temperature fraction for different samples). I strongly suggest instead plotting each sample as a line in ^{14}C / $\delta^{13}C$ vs. temperature (or, better, activation energy) space so that the reader can follow trends for any given sample.

Additionally, for Fig. 3, the authors simply omit the 650-800 °C data. Only later in the discussion did I realize this is because the authors attribute these enriched values to carbonate contribution. However, there is no mention of this in the figure---a seemingly dishonest omission. I strongly suggest the authors include these data in the figure---they are valid data after all---and describe why they behave differently from the rest.

Finally, there are several instances where statements and interpretations seem to be in direct conflict (e.g., L132-135: “There were no significant trends with depth … Visually, for both T1 and T3 the size of the second major peak…”; similarly repeated on L210-211). Either the size of the peaks changes between samples, or it doesn’t---if it is statistically insignificant, then any visual differences are moot and should not be interpreted. However, I suspect some of this statistical insnignificance is due to the particular method that the authors normalized their thermograms (see above comment).