Response to reviewers

**Reviewer #2**

Evaluations  
Recommendation (Required): Return to author for major revisions  
Significant: The paper has some unclear or incomplete reasoning but will likely be a significant contribution with revision and clarification.  
Supported: Mostly yes, but some further information and/or data are needed.  
Referencing: Mostly yes, but some additions are necessary.  
Quality: The organization of the manuscript and presentation of the data and results need some improvement.  
Data: Yes  
Accurate Key Points: Yes

General comments:

1) I appreciate that the authors have clearly stated their hypotheses in the introduction.

However, the introduction needs a clearer structure and progression of information. For example, it is not until lines 134-137 where the study intent is indicated, and it is not clear until this point what the authors are interested in studying. Most of the necessary information is here but needs to be refocused. I recommend the authors shift their existing paragraphs into this general structure: carbon cycling, air-drying and rewetting, soil archives, radiocarbon dynamics, study objective and hypotheses. As I have already mentioned, all of these components are already in the introduction, but have been integrated in a way that is confusing in present form.

We agree with this assessment. We have revised the introduction to emphasize and refine the main goals of the study. The focus of the study is introduced much sooner now: we introduce radiocarbon incubations in the third sentence, while the primary study objective (measuring ∆14C-CO2 in incubations of archived soils) is presented in the third paragraph.

We reorganized the introduction mostly following the reviewer’s suggested structure. Specifically, the order of topics is now: 1) carbon cycling, 2) soil archives, 4) radiocarbon dynamics (bomb-C), 4) air-drying and rewetting, 5) research questions and overview of key findings. The original structure emphasized air-drying and rewetting more than we felt was appropriate, given that we designed the study to assess bias in ∆14C-CO2 due to archiving, so we reorganized accordingly.

2) The methods are an integral section of the paper and need some more clarification. I found myself constantly flipping back between Table 2 and Section 2.3 to understand the different treatments for each of the samples and neither the table nor the text appear complete. For example, all of the information for Experiment 3 samples is included in a large supplementary table. The summarized information for all three experiments needs to be included in the manuscript.

There are several points where the motivation in the methods is unclear. For example, why change the equilibrium respiration for the different experiments (~line 247)? I worry

that the authors created too many variables to be able to truly compare results from the different experiments without a proper normalization, which I understand can be difficult

(e.g. lines 245-266). For experiment 3, it is unclear why the methods for experiments 1 and 2 could not be applied. Maybe I missed something here. These sections can be greatly strengthened by adding more justification.

We reorganized the methods section in response this comment as well as a comment from Reviewer 3. We now present the methods by describing each experiment in turn. We think that the expanded text and new organizational structure provides clarification why we needed to treat the samples from Experiment 3 differently, and why our approach was justified in the context of our research goals. However, we also expanded the text and Tables 1, 2 and 3 accordingly to incorporate data from Experiment 3 into the main manuscript, rather than relegate the data to the supplemental information.

We acknowledge the concern about comparisons across experiments. However, the primary goal of this study is to assess bias in ∆14C-CO2 in incubations of air-dried and rewet samples with and without storage in comparison to incubations of samples that have never been air-dried. Unlike experiments focused on respiration rates, where equilibrium respiration is important, the objective of radiocarbon incubations is to respire enough CO2 to measure 14C while not respiring so much C that the age structure of respired CO2 is altered. Accordingly, we have removed the phrase “equilibrium respiration” as we recognize that we did not measure equilibrium respiration in all experiments or for all samples. We believe that the differences in ∆14C-CO2 observed between control and treatment samples are comparable across all three of our experiments because we have controlled the amount of CO2 respired (per g of soil C) between control and treatment samples in all three experiments, and also kept the soil moisture the same between control and treatment samples.

3) Section 3.4 relies heavily on interpretations from Figure 4 and the finding that storage duration does not affect Δ14C is a critical conclusion for the paper. I generally agree with the authors’ interpretations. However, they argue for a significant difference between control and treatment samples with differing storage durations for Oak Ridge. I do not see this same result. If you take the average of the samples/sites with consideration of error, there really is not a large statistical difference between the samples stored for around 5 years and those stored around 14 years. I recommend the authors instead focus on general observations. For example, one of the 14-year Oak Ridge sites has the lowest treatment-control value, etc.

We agree that this analysis was confusing in the context of the results as presented. We have revised the text to remove discussion of statistical significance. We have also changed the graphical display to show standard deviations rather than 2x standard error. The two points shown in pink for the Oak Ridge sites at each storage duration are from different sites with different levels of exposure to the 14C labeling plume: Tennessee Valley and Walker Ridge (Cisneros et al., 2006). However, we do provide a more in-depth analysis of these data in the supplemental information. Comparing the interaction of storage duration and treatment in a linear mixed model with sample ID as a random intercept term shows that the interaction is significant when the Oak Ridge samples are included in the overall dataset. This significance disappears when the Oak Ridge samples are removed from the model. We believe it is reasonable to remove these samples from the model as they represent a different population than the other samples and provide unfair leverage since they happen to have been stored for the longest time. However, we have removed the discussion of the mixed modeling approach from the main text in the interest of clarity. Please see the supplemental information for the details of these methods and an explanation of the results.

4) These next few comments focus primarily on the discussion.

a. A few portions of the discussion are reiterations of the results (e.g. lines 610-617;

627-629). The authors should review to make sure the interpretations and applications of the results are the focus.

We have cut down the results section substantially in order to remove text that had more of a discussion tone and focus on the key results. Additionally, we have also removed text from the discussion when we thought it was redundant with the results. We have also reorganized the discussion section to emphasize the key findings and structured it around the research questions we pose in the introduction.

There are a few areas I really wish the authors expanded upon, particularly regarding slow and fast pool dynamics. In particular, I am curious how likely the scenario of pool reversal is on lines 577-581. Without more discussion in this section (4.2), it feels like a large leap to conclude that air-drying and rewetting results in different rates of carbon cycling between forests and grasslands and mobilizing different pools (lines 589-592).

We did not intend to convey this result. We have revised the discussion to clarify that we think we have observed is that the air-drying and rewetting treatment results in a shift in observed ∆14C-CO2 of different magnitude due to different *inherent* rates of soil C cycling in these two different ecosystem types (or due to vegetative lag effects in forests, as discussed). We do not infer a causal relationship between air-drying and rewetting effects and soil carbon cycling rates. Rather, air-drying and rewetting makes soil organic matter of different ages available to microbial community in laboratory incubations of forest and grassland soils, with the result that grassland soils show a greater magnitude of difference between control and treatment sample incubations.

The relevant section of the discussion section now reads:

“We hypothesize that the disturbances associated with air-drying and rewetting is from mobilization of carbon from more slowly cycling pools. For soils sampled at different times and given the trajectories of ∆14C in slow and fast cycling soil carbon pools over time, we might predict different responses. The importance of the year of sampling and carbon dynamics is illustrated using a conceptual model developed by Schrumpf & Kaiser (2015) for forest sites in the Hainich-Dün (Fig. 6).

Comparing the data from the Hainich-Dün forest sites with model projectsion of the trajectories for fast, slow and respired ∆14C (Figure 6) indicates they are consistent with mobilization of carbon from the slow C pool after drying and rewetting. Following treatment, ∆14C-CO2 of respired CO2 (black points) shifts towards the dashed blue line that shows the trajectory of slow pool ∆14C over time, indicating increased contribution to respiration from this pool. Due to the crossing of the slow and fast (magenta) pool curves in 2015, an increased contribution of the slow pool to respiration following treatment leads to relative depletion of ∆14C-CO2 in 2011, but relative enrichment of ∆14C-CO2 in 2019. Thus, depending on the sampling year, the bias in ∆14C-CO2 introduced by air drying and rewetting could be either higher or lower relative to a sample incubated without air-drying.”

I would also like to see the authors expand the discussion on extracellular carbon on lines 641-643, if possible.

Unfortunately we did not design the experiment to investigate this mechanism, so while we think this is an important line of inquiry, we have cut out this speculation on the role of extracellular carbon in driving the observed treatment effects.

I do not believe the discussion will be significantly longer with the suggestions above once the redundant portions from the results are removed/condensed.

Agreed. We have implemented the recommended changes and the overall word count of the manuscript has been substantially reduced.

b. On lines 544-545, the authors state that the decomposition rates they selected were arbitrary. They use this rate to create a simple model shown in Figure 6.

Either sound justification for the selected rates need to be provided, or the authors need to be much more general in their interpretations. If the decomposition rate is arbitrary, then only trends can be discussed in lines 550-555, and not specifics as the authors have currently written. On line 566, the authors state that the carbon pool becomes enriched in Δ14C in the mid-1970s based on Figure 6. This is similar for line 569. Again, if the decomposition rates are arbitrary, the year is not accurate/important. It’s fine to mention years for the figure, but it needs to be clearly stated that these are not representative (or if they are representative, more justification needs to be provided).

We have clarified the basis for the conceptual model that we present. The model we used was originally fit with data collected at the same site (Hainich-Dün forest) in an earlier study (Schrumpf and Kaiser, 2015). We had initially made minor adjustment to the model to illustrate the potential mechanisms of slow vs. fast pool mobilization. However, when reevaluating this approach in response to the reviewer comments, we decided it would be simpler and more powerful to use the model as it was developed by Schrumpf et al. (2015). The details of the model development are now given in the method section (§2.7):

“We developed a conceptual model for the forested sites from a single region, Hainich-Dün (Central Germany 2), to illustrate potential sources for the carbon respired following the air-drying and rewetting treatments imposed in this study. We implemented a two-pool parallel model with inputs partitioned between slow and fast cycling soil C pools, and no transfers between pools, using the Soil R package (Sierra et al., 2014). In an earlier study, Schrumpf & Kaiser (2015) estimated intrinstic decomposition rates and pool sizes for empirically defined soil C pools using a density fractionation procedure for sites in the same region. We approximated the intrinsic decomposition rates for the fast and slow pools of our model using Schrumpf & Kaiser (2015)’s mean estimates for the free light fraction and the heavy fraction from the 0-5 cm depth increment (4-1 y and 115-1 y for the fast and slow pools, respectively). Schrumpf & Kaiser (2015) found that 10 percent of the carbon in the 0-5 cm depth layer was in the free light fraction. We used this fraction for the partitioning between the fast and slow pools, under the assumption that the free light fraction corresponds to the fast pool. Following Schrumpf & Kaiser (2015), we assumed a lag time of 8 y for inputs.”

I understand that a complete model is not within the scope of this study. However, the authors base much of the discussion on a model with arbitrary values. Moving forward, I recommend the authors provide clear justification for their model inputs and dial back their interpretations to focus more on trends and general pool behavior. The current text on lines 570-572 is a great example.

We have followed this suggestion. We have excised the detailed explanation of the conceptual model curves and now focus only on the time period relevant for our observations (~2010-2020). See the quoted text in response to your question 4)a.

5) The figures generally look great and the abstract and plain language summary read well.

Thanks.

Line comments:

Line 74: Use of attack is odd; define artifacts. We have removed this sentence from the introduction

Lines 76-77: See comment above. We have removed the mention of artifacts.

Line 84: I would remove this topic sentence. Removed.

Lines 91-92: Can this be expanded? It is not clear what is meant. We have clarified this with a worked example and rewritten the paragraph: “A critical issue with interpreting bomb-C radiocarbon is that there are two points in time at which the Δ14C signature of atmospheric CO2 is identical, due to the curvature of the bomb-C peak. This means observations of Δ14C from a single point in time can be fit to models with different rates of intrinsic decomposition. Trumbore (2000) gives the example of a two independent, homogenous pools of soil carbon, one with an intrinsic decomposition rate (*k*) of 6.6 years and the second with *k* = 50 years, both of which would have a Δ14C of 166‰ in 1996. Observations of Δ14C-CO2 measured in incubations of archived soils could help solve this problem by enabling the contruction of a Δ14C-CO2 time series. The trajectory of ∆14C in a soil carbon pool turning over every 6.6 years is quite different from one with an intrinsic decomposition rate of 50 years (Baisden et al., 2013), a powerful additional constraint for model parameterization.”

Line 151: This study did not really “assess the feasibility of measuring Δ14C”, correct? I

understood it as quantifying post-collection alteration. Yes, this is an important distinction. We have revised this accordingly, vis. “Obtaining ∆14C-CO2 measurements from incubations of archived soils would be a valuable tool for further constraining and improving soil carbon models, but first the possible effects of air-drying and rewetting, as well as the effect of storage duration, must be assessed.”

Line 173: Delete the second “within the”. Done.

Lines 201 and 205: What are the number of control samples for the three

experiments/treatments? Can these numbers be included? These numbers have been added, vis. “We then selected soils from three grassland plots (50 m by 50 m) and three forest plots (100 m by 100 m) in each of the two geographic regions (n total = 12 sites)...”. See also Table 1.

Line 235: Was there a target range for the soil carbon content? We have clarified the text after conferring with the original investigators. In short, no, there was not a precise range, rather respiration rates were estimated from earlier work at the sites, and the mass of soil was adjusted accordingly so that CO2 production would be enough to measure ∆14C-CO2 while not reaching microbially toxic levels in the incubation jars. “The mass of soil used for control-1 incubations ranged from 45 g to 75 g (air-dry equivalent) based on estimated respiration rates from previous work at the sites. Soil masses were adjusted to ensure that enough CO2 would be respired to measure ∆14C-CO2 (> 0.5 mg) following the second enclosure period while at the same time preventing excessive CO2 build-up as this has been shown to negatively impact heterotrophic respiration (MacFayden 1973; Santruckova and Simek 1994).”

Line 238: How did you adjust the soil moisture to 60%? Clarification added, “We moistened the soil from the top by means of a perforated luerlock cap attached to a 10 ml syringe that emitted water in small droplets for minimal disturbance.”

Lines 245-247: Which study did the authors reference for the incubation duration? The original investigators did not reference a previous study for the incubation duration, rather the duration was chosen to allow samples across a range of potential respiration rates to accumulate enough CO2 to measure ∆14C-CO2 while avoiding microbially toxic CO2 concentrations. See above response to the comment on Line 235.

Line 266: Why not measure equilibrium for all samples? The phrasing “equilibrium respiration period” has been replaced with “second enclosure period” throughout the manuscript. The goal was not to measure equilibrium respiration, rather to isolate the CO2 released during the rewetting pulse from CO2 released during a second enclosure period. While equilibrium respiration would be interesting to measure, the intention of this study was to assess whether air-drying and rewetting, with and without storage, would change the ∆14C-CO2 measured in soil incubations, which we do not believe requires the quantification of equilibrium respiration. We think the phrase “second enclosure period” is more accurate and hope that this eliminates confusion regarding equilibrium respiration. As mentioned above (comment #2), the objective of radiocarbon incubations is to respire enough CO2 to measure 14C while not respiring so much C that the age structure of respired CO2 is altered. Accordingly, equilibrium respiration is not a prerequisite.

Line 268: As closely to what? This section has been rewritten to clarify why the storage duration incubations were performed in the way that they were, and why the protocol for Experiment 3 incubations differed from that of Experiments 1 and 2. “Control-3 incubation soil mass, replication, temperature, and adjusted moisture varied according to the objectives of the original investigators (Table 2). Soil mass and replication of corresponding storage duration treatment sample incubations varied by the amount of soil material available. We kept the soil moisture the same between paired control-3 and storage duration treatment incubations. However, all storage duration treatment incubations were conducted at 20ºC for simplicity, as while temperature has known effects on respiration rates, it has been shown that it does not affect ∆14C-CO2 (Vaughn and Torn, 2019).

We did not have information on either the duration of or the amount of CO2 respired during the rewetting period for all of the control-3 samples. Rather than impose a first enclosure period with an arbitrary duration, we decided to incubate the storage duration treatment samples for a single enclosure period beginning immediately after rewetting. We felt this was justified as we did not observe significant differences between first and second enclosure period ∆14C-CO2 in the first two experiments (Results 3.2). We allowed respiration in the storage duration treatment samples to proceed until the same amount of CO2 had been respired per g of soil C as in the second enclosure period of the corresponding control-3 sample incubations.”

Line 280: How was the moisture adjustment determined? We have clarified this in the text: “An additional 10g aliquot was removed from the air-dried sample and further dried at 105ºC in order to determine water holding capacity (WHC). Briefly, the tip was removed from a 50 ml centrifuge tube and covered with a fine mesh (<50 µm). The tube was filled with soil and placed mesh-side down overnight in a glass dish filled with a mass of water equal to twice that of the mass of soil. The following day the tube was moved to a second glass dish filled with sand and allowed to drain for 30 minutes before weighing again (wet mass). WHC is defined as the difference between the wet mass and the oven-dry mass over the oven-dry mass.”

Line 297: Can a sample number be listed instead of “a few”? We have revised this to provide the sample numbers explicitly: “We conducted the majority (n = 16) of the Experiment 3 storage duration treatment incubations in 2018 at the Max Planck Institute for Biogeochemistry (MPI-BGC) but the remainder (n = 12) of the treatment sample incubations were performed in 2009 at the University of California Irvine (UCI) (SI Table 1).”

Lines 319-322: Important information, but not methods. Agreed, this has been removed.

Line 335: A familywise error correction would be helpful to make sure the authors are not committing a type-1 error. We did not actually make multiple comparisons using the t-test approach, but rather within experiments and within ecosystems. However, we did not incorporate the laboratory error properly in the initial analysis, so we redid this analysis in order to rectify this. We first quantified the analytical error associated with the radiocarbon incubation method by calculating the mean of the variance measured among replicates for all samples that were replicated. We assumed this mean variance for those samples that were not run in replicate so we could more accurately calculate the error of the difference between treatment and control means for each sample. We have revised the description of the methods to clarify this. Additionally, we evaluated contrasts across experiments and ecosystem types using a mixed modeling approach. For the mixed modeling we were able to use the raw data without the assumed variance. This was possible because we set sample ID as a random intercept term in order to account for the different number of measurements made on each sample (e.g. replication in treatment samples but not control-1 samples in Experiment 1). However the description of this secondary approach and the subsequent results have been moved to supplementary information in the interest of clarity, since they were similar to the simpler approach described in the main text.

Line 340: What is meant by “field-moist”? This qualification has been deleted. Additionally, we have removed mention of statistical significance here. The revised sentence reads, “Maximum respiration rates were more than twice as high in grassland soils than in forest soils for air-dry/rewet + storage treatment samples in Experiment 1 (Fig. 1a), but were similar across ecosystem types for the air-dry/rewet treatment samples in Experiment 2 (Fig. 1b).”

Lines 361-362: Are those really peaks? We have changed the wording, “the magnitude and timing of maximum respiration rates diverged among experiments and between grassland and forest soils (Fig. 1).”

Lines 369-370: Maybe I’m missing something because the chart is in days, but I don’t see peaks at the interval mentioned. Agreed, wording has been changed. We have also changed the units in the text to days to agree with the figure.

Line 371: Why is this in the supplements? Is it not possible to include experiment 3 in Figure 1? We have added text to clarify why the respiration rates measured in Experiment 3 cannot be not meaningfully interpreted, and therefore are not shown in the main text, “Respiration rates for Experiment 3 samples are shown in Supplementary Figure 1, but CO2 flux rates cannot be meaningfully interpreted given the differences in incubation temperature, approach to the rewetting pulse, and the wide variation in CO2 measurement frequency among samples.”

Lines 386 & 391: Which tests were used to determine significance? What are the values?

Information is missing. We have removed the statements of significant differences between respiration rates as we realized it was not particularly relevant in the context of our study, and instead focused on the general trends that are self-evident in the data. However, in response to this comment we have expanded the statistical analysis section of the methods substantially in order to clarify our approach. We have also added the relevant confidence intervals in the results section, where appropriate, in order to make explicit which differences we are considering statistically significant. Briefly, we determined mean differences between control and treatment samples within experiments and ecosystem types to be significant if the confidence interval around the mean difference excluded zero. These tests were only run for ∆14C-CO2 and δ13C-CO2. However, given the relatively small sample sizes in this study, we chose to deemphasize statistical significance when we revised the text. Instead we focus on the absolute values of the mean differences and associated variances observed in the context of the main goal of the study, which is to assess the bias introduced by air-drying/rewetting and storage on ∆14C-CO2 and the implications for constraining soil C models.

Line 411: I don’t agree that this is a significant depletion since the values do have overlap. This section has been revised to deemphasize this as it is misleading given the rest of the data. For the effect of the Oak Ridge samples on a mixed model regression assessing the interaction of storage duration and treatment on ∆14C-CO2 see the supplementary information. When included in the model, the interaction is significant, but it is a leveraging effect that is conflated with the increased difference observed in these samples due to artificial labeling and the fact that these samples happened to have been stored for the longest amount of time. When the Oak Ridge samples are removed from the dataset, this interaction is no longer significant. However, as this discussion is not germane to the primary narrative of the study we have moved the mixed modeling and the discussion of these subtleties to the supplementary information.

Line 415: Typo, should read “not measured”. Fixed.

Line 502: I interpreted this as a slightly lower mean for the forested sites. Yes, this section has been revised for clarity.

Lines 521-524: This is confusing and I’m not sure what is meant. Agreed. We have revised this section for clarity.

Line 524: A significant change relative to what? The references to significance have been removed except for where the appropriate statistic or comparison has been clarified.

Line 533: Shouldn’t this be incubation practice, not technique? Replaced “technique” with “approach”.

Line 550: Maybe I’m missing something, but I don’t see the shift towards the slow pool curve. It looks like the respiration line to me. We have revised the figure to show the curves from the two scenarios and removed the modeled control curve for clarity.

Lines 605-608: How would you determine if the pool curves did indeed cross? Added the following sentence, “This could be tested by comparing the ∆14C of empirically defined “fast” and “slow” carbon pools over this time period relative to observed ∆14C-CO2, e.g. free light particulate organic matter and mineral associated organic matter. However, this is beyond the scope of this study.”

Lines 622-624: This sentence is hard to understand. We have changed the first two sentences of this paragraph to, “We hypothesized that the CO2 released immediately following moisture adjustment, i.e. the rewetting pulse, would be older than the CO2 released during the second enclosure period. However, in contrast to our expectations, we did not find a significant difference in ∆14C-CO2 between these two respiration periods.”

Line 645: I would just start this section with “The consistent enrichment…”. Changed accordingly.

Line 646: Instead of microbes, is it possible that the enrichment could be due to decomposition? Yes, this is possible, although unlikely given that little change in soil C is observed when samples are carefully air-dried and properly stored. We have added this explanation and expanded the discussion of this scenario.

Line 654: “a process [that] has also…” This has been fixed and the section revised for clarity.

Lines 665-669: Is there actually evidence at the study sites for mineral-carbon release? Are there limestone/calcite/dolomite beds? We were not referring to inorganic carbon release, but rather mineral-associated organic carbon. This has been clarified.

Line 680: I disagree that the Oak Ridge data provide \*strong\* evidence of different carbon pools. A couple of typos here too: “data provides…”; “fixed from the atmosphere in over the…”. Agreed. We have revised this section to expand on this line of reasoning and removed the phrase “strong evidence”.

Line 689: This explanation seems fine, but I’m not sure it explains the variability for the rest of the samples. We have rewritten this section for clarity and removed most of this text, as it is redundant with previous discussion.

Line 709: Sorry if this is a silly question, but what is meant by protected carbon in this context? Agreed, this is not clear. We have removed these sentences.

Lines 734-737: I wonder if a correction could also be appropriate here. While a correction could be appropriate in this case since we observed ∆14C-CO2 of both field-moist and air-dried soils, this would not be possible for in many situations. Our interpretation of the data suggests that the inference of a correction for the air-drying and rewetting bias would be have to be made on a case-by-case basis. Absent more information about the drivers of the air-drying and rewetting effect, we do not feel that is appropriate to apply a quantitative correction. Instead, as we have tried to convey, we think our results suggest that the carbon respired in the air-dried and rewet soils studied here appears to be dominated by C fixed from the atmosphere between 10 and 20 years prior to sample collection. However soils with very different mineral assemblages or from very different climates or ecosystem types may respond differently. We have added a sentence with this caveat in the conclusion.

Figure 1: All results are discussed relative to hours. It would be more helpful for the x-axis to include hours as well. We changed the text in the results to units of days for consistency with the figure.

Figure 3: Can the different treatments be shown somehow? We have added information about the experiment (gray shading) to this figure.

Figure 4: The caption says all three experiments, but I believe only experiments 1 and 3 are shown. You are correct, thank you for pointing this out. We have updated the caption accordingly.

Figure 5: Can the grey line be made black instead? Can the x-axis include 2011 as well? The last line of the caption should read “radiocarbon data are…”. We changed the atmospheric 14C curve to black, modified the x-axis labels to include 2011, and fixed the grammatical error in the caption.

Table 2: I really like this table, but it needs a little more detail. I mention this previously, but instead of saying “various” for experiment 3, can the authors include ranges with standard deviations? Good suggestion. We have replaced “various” with the ranges.

**Reviewer #3**

Evaluations  
Recommendation (Required): Return to author for minor revisions  
Significant: The paper has some unclear or incomplete reasoning but will likely be a significant contribution with revision and clarification.  
Supported: Yes  
Referencing: Yes  
Quality: The organization of the manuscript and presentation of the data and results need some improvement.  
Data: Yes  
Accurate Key Points: Yes

Radiocarbon has been a critical tool for exploring the C cycle, particularly a tool for exploring long-term dynamics; bomb carbon explores decadal scales, while traditional radioactive decay C-dating explores century scales. A challenge though, has been methodological: how to analyze soil samples to best evaluate the 14C. This paper addresses a few issues, notably how to use archived soil samples to look at changes in C pools over multi-year time scales.   
  
What I found, however, was that the paper struggles with whether it is a methods paper, asking about the effect of storage and rewetting on the measured ∆14C, and whether it is a "science" paper exploring the nature of the internal dynamics of soil C as revealed by the 14C data. The paper reads as a methods paper into the results, but begins to shift in the discussion. The hypotheses state are introduced by "hypotheses regarding the potential effects of air-drying, moisture adjustment, and storage duration on Δ14C-CO2 observed in laboratory soil incubations" which firmly set the methodological focus. And part of the conclusions supports this as well:   
"Overall, the results of this study suggest that measuring the Δ14C of respired CO2 in laboratory incubations of archived soils is a promising technique for improving quantitative interpretation of soil C dynamics and can provide a strong constraint for soil C models in the future."  
  
But much of the discussion is more conceptual and comes across as the science paper evaluating the nature of C dynamics in soil, and how grassland soils behave differently than forest soils. That's not a terrible problem, but it made it harder for me as a reader to figure out what the central message was. There's more text there than we need for the methods story, but the focus is a little off to prepare readers for the discussion of mechanisms, which is a little incomplete. For example, in comparing forest and grassland, and the different balance of stabilized, pre-aged, or fresh inputs, the authors don't discuss some aspects that might also play in-grasslands are often found on soils that have more clay development and accumulation, which can stabilize C.

The authors thank the reviewer for the thoughtful and helpful comments. As was clearly apparent from the initial version of the manuscript, this study started out purely methodological, but evolved to be more conceptual as we collected more data to try to explain our initial results. The initial goal of the study was to assess the feasibility of constructing time series of ∆14C-CO2 from archived soils as a way to constrain soil carbon models; we did not design the study to test truly mechanistic hypotheses about possible drying and rewetting effects. We have revised the introductory section accordingly in order to clarify the motivation of the study and remove extraneous text about mechanisms of air-drying and rewetting. Additionally, we have revised the discussion section to focus on interpreting our results in the context of the target application, i.e. constraining soil carbon models.

The description of the experiments is also a little unclear-the authors offer 3 hypotheses, but then there are 3 experiments but the linkage between these is unclear, and nowhere are each of the experiments clearly described, one at a time. The first paragraph of the methods tells us what the experiments consider, but not what they actually do. But the methods cut across the experiments: field sampling, sample prep, incubations. But we get the info for the three experiments simultaneously and have to sort them ourselves. A more natural approach (for a reader) would be to describe each experiment in turn.

We agree that the initial structure of the methods section was confusing. Reviewer 2 also highlighted this. We have followed the suggested approach and now describe each experiment in turn. We have also revised the end of the introduction to focus on research questions that correspond to each of the experiments. The hypotheses are described separately.   
  
Beyond that, I have just some minor issues.  
  
Line 627: "These results suggest that the change in substrate availability initiated by air-drying and rewetting may not be limited to the rewetting pulse, and may persist throughout the incubation." This contradicts most work on dry/rewet cycles that show that the resources that fuel the rewetting pulse are distinctly different than those that fuel the basal activity. Maybe they share ages, but are likely to be chemically distinct.

Agreed, and this is a good point. We have added, “There is a large body of literature that provides evidence for different chemistry of the substrates fueling the rewetting pulse compared to that of the substrates fueling basal respiration (cite, cite, cite, etc.). However, as other recent work has shown, persistence of soil organic matter is not soley due to chemistry (cite, cite, cite, etc.), so this similarity in ∆14C across substrates utilized in the rewetting pulse and the second enclosure period, which likely diverge in chemistry, is in line with the modern paradigm (Lehmann and Kleber, 2015).”

645: "Alternatively, the consistent enrichment in δ13C-CO2 seen following both the air-dry/rewet + 645 storage treatment and the air-dry/rewet treatment" How do you start a new section (4.4.3) with   
"alternatively"? The section header says, "this is a new thought;" the word says "This is a continuation of the previous thought." Good point. This has been revised.  
  
664: "more mobilization of mineral-associated carbon in forest soils than in grassland soils following treatment" This makes some sense-forest soils usually are more poorly developed with less clay. Hence mineral associated OM might be less tightly held. But little is ever discussed about the nature of mineral associations or what might make them stronger or weaker. This is a good point. Owing to the limited data at our disposal regarding mechanistic explanations, we have also limited the speculation about these differences.