

We would like to thank the referee for taking the time to read and comment on our manuscript, which will help us improve its clarity in a revised version.

Answers to individual points raised follow below.

Referee 1

The manuscript reports on experiments characterizing mineral associate organic matter using a comparison of "extraction" by NaOH-NaF versus "oxidation" by H₂O₂. The treatments were coupled with ¹³C-NMR and ¹⁴C dating. The experiments are well designed according to current paradigms of SOM stabilization, but the manuscript requires substantial revision before it might be acceptable for publication.

A revised manuscript will be prepared based on the recommendations and suggestions provided by the two referees.

There are a large number of instances throughout the manuscript where diction and grammar are awkward or incorrect. The manuscript should be deeply reviewed by a native English writer to revise these subtleties.

As recommended, we will adopt the semantic changes suggested by the two referees, and the two native English speaking co-authors of the manuscript will double-check the overall language again for awkward wording and syntax flaws.

The end of the introduction should be restructured to move ln96-104 to the end of the section, where it will serve as a segue to the methods section. That is, I recommend an introduction with the structure: background, problem statement, hypotheses, ways to test them.

We will transfer the description of the methods used to test the hypotheses to the end of the introduction to obtain the structure recommended by the referee.

The hypotheses listed are not hypotheses *sensu stricto* because they are not testable in the strictest way. That is, they cannot be answered with a simple "yes" or "no". Several hypotheses have multiple conditions or clauses that should be broken into several subhypotheses. I am not necessarily a purist when it comes to these formulations, but when I know this, I typically replace the word "hypothesis", which can be reserved for strict statistical or logical uses, with the word "expectation". All this being said, I do recommend trying to break up the 5 bullets into expectations and outcomes. That is, it appears that some of these so-called hypotheses are actually expectations, which if true, other conditions would also be true. Others are more contradictory, where if not true then... A clearer, more explicit and deliberate structure for all these expectations will result in better structured results and discussion sections.

We agree having several conditions in one hypothesis hampers simple yes/no answers to whether or not a hypothesis is supported by the data, as the answer could depend on the conditions.

We suggest rephrasing and simplifying the "hypothesis" in the following way:

Based on the literature review we expect that:

- (1) extraction in NaF/NaOH releases a weaker bound fraction of total MOC, which is younger than the stronger bound, probably better stabilized residue fraction.
- (2) the proportion of NaF/NaOH-extractable MOC decreases with increasing contents of pedogenic oxides, which form strong bonds with OC.
- (3) the proportion of NaF/NaOH-extractable MOC declines from topsoils to subsoils due to declining OC loading of minerals.
- (4) MOC of soil samples with higher portions of extractable carbon should be younger than MOC with more extraction resistant OC.
- (5) the chemical composition of extractable MOC varies between study sites due to differences in vegetation composition, and thus litter chemistry.
- (6) the chemical composition of extractable MOC changes with soil depths due to declining contributions of plant and increasing contributions of microbial derived OC.
- (7) MOC should be less prone to desorption and accordingly of older ^{14}C age when organic molecules forming MOC have many carboxyl groups that enable strong bonds with mineral surfaces.
- (8) the strong oxidizing agent H_2O_2 removes more OC from MOC than NaF/NaOH.
- (9) oxidizable and non-oxidizable OC should both be older than the extractable and non-extractable OC fractions if the harsher oxidation treatment removes more, and thus presumably better stabilized OC from mineral surfaces, leaving OC residues of even older ^{14}C age behind.

A few minor points in the methods section:

In141: filters are normally described at least by their pore-size and sometimes their diameter, not their diameter only.

The diameter was 1.6 μm and this information will be added to the revised version.

In161-162: The sentence on carbonates is unnecessary and likely the result of copy-paste text because In158-159 already state that all soils were carbonate-free.

The referee is right that carbonate-free samples were chosen, and the sentence will be deleted in the revised version.

In some instances the order in which results are circuitous and confusing. Perhaps use the "hypotheses" or methods section as road maps for ordering the presentation of results.

We assume that mixing of ideas on the effects of NaF-NaOH and H_2O_2 was possibly confusing. Therefore we suggest to restructure the results section as follows: First, presentation of the extraction efficiency of NaF-NaOH, followed by the chemical composition of the extracts (NMR), and then of the ^{14}C data. Thereafter, presentation of the oxidation efficiency and ^{14}C contents of H_2O_2 treated samples. Finally, a comparison of results of the two treatments will be presented.

I like how the discussion section is structured. It places the results in a clear context. However, I'm not convinced that the results are "surprising", and I find the overall

interpretation to be a little off the mark. I did not find it surprising that the proportions (vs."portions" which is incorrectly used throughout the manuscript) were consistent across most soils. This has been previously observed for both H₂O₂ (eg, Plante 2014 EJSS) and acid hydrolysis (eg, Paul 2006 SSSAJ). This is one of the problems with chemical extractions: they rarely demonstrate the expected trends in inferred stability. A similar lack in expected trends has also been frequently observed in ¹⁴C dates.

The surprising result was not the lack of a trend in H₂O₂ oxidation but the uniform extractability in NaF-NaOH. While the two references mentioned by the referee (Plante et al. 2004?) indeed found similar proportions of soil OC removed by H₂O₂ treatment and acid hydrolyses, others did not. For example, Eusterhues et al. (2005 Organic Geochemistry) showed a strong depth dependence of H₂O₂- or Na₂S₂O₈-oxidizable proportions of bulk OC (between 5 and 58% of OC resisted oxidation with H₂O₂) and also Kleber et al (2005 EJSS) found a range of 28-87% for oxidation-resistant OC across different forest subsoils. After all, while studies on OC oxidation and hydrolysis were frequently done before and address rather the inherent chemical stability of soil OM, the focus of our study was on using OC extraction into NaF-NaOH as an indicator of the stability of the bonding between OC molecules and mineral surfaces (which is also why mineral associated and not bulk soil OC was studied). The controls on OC mobilization are therefore not necessarily the same and can be expected to be closer linked to mineral composition and OC loading of minerals, as indicated by previous results from the literature. The same applies to ¹⁴C data. Therefore, the focus of the discussion was also on the NaF-NaOH results, while little emphasis was placed on the results of oxidation resistance since not specifically surprising.

We will replace "portion" by "proportion" in the revised version of the manuscript.

So, given that a substantial proportion of the results from this study did not meet expectations, I would strongly recommend reframing the manuscript. It might be much more compelling to more specifically spell out what the conceptual framework (paradigm) is that leads to the expectations outlined. The goal of the discussion would then be to describe where the problems are with either the assumptions etc. in the conceptual framework, or in the methods used to test them. The manuscript currently tries to address the former, but not necessarily the latter. Are NaOH-NaF and H₂O₂ appropriate tools for probing SOM stability? OR is our conception of SOM stability incorrect?

As the referee pointed out, we had a certain concept about MOC formation, its stability and potential drivers in mind when we started the experiment as outlined in the "expectations". Many of them were not supported by the data, which can indeed be due to a combination of wrong methods used or because our conceptual understanding needs further refinement. In any case, changing the expectations a posteriori to make them meet the results is no reasonable way to achieve scientific progress. Instead, the discussion should be used to identify where our conceptual understanding is not in agreement with the results, derive hypotheses for the observed mismatch including potential limitations of the experimental approach, and eventually draw a modified version of the original concept and come up with ideas for further studies. Therefore one could also argue that it is useful if an expectation is not met, because it means that we can learn something from the study that we did not know before.

Nevertheless, one result of our study was that a large portion of MOC could be more homogenous in ^{14}C contents than expected, while only a very small portion has very old ages. Accordingly, different chemical fractionation schemes always remove OC from the same continuum, leaving increasingly old OC behind. Therefore, chemical extractions seem to be not able to isolate distinct homogenous fractions, and other approaches are needed to explain average ^{14}C values and age distributions of MOC as pointed out in the conclusions.

The match between expectations and results that I am referring to is well illustrated in the diction of the subheader in In344. The use of the word "Missing" suggests it was expected a priori. A more objective and unbiased approach would be to refer to it is "lack of".

A negative statement in the headline will probably always hint at an unexpected result. Nevertheless, we will replace “missing” by “no”, which is more neutral than “lack of”, which, at least to our understanding, still indicates that something was expected to be there.

I also found it unsurprising that the chemistry of extracted OC was similar across samples and dominated by polar molecules (eg, alkyl). In essence, the experiment demonstrated the solubility of a polar fraction of OM in a highly polar solution. It would not be a reasonable expectation to see non-polar OM (eg, aryl) in such a polar solution, or vice-versa.

We agree that we were extracting the samples in a polar (aqueous) solution and that accordingly most of the extracted molecules are dissolvable in water – which is indeed not surprising given that they were probably transported and transferred to mineral surfaces via the aqueous soil solution. In general, alkyl-C and aromatic C (aryl-C) moieties would be non-polar, while carboxyl groups are polar. In that respect, we could state that it is rather surprising that the extracted fraction was dominated by alkyl-C. However, whether or not a molecule is soluble in water– similar to aryl-C- dependents on the type and number of functional groups attached. It is not possible to infer the solubility or polarity of extracted molecules from their NMR spectra, as these give only information on bond types of elements but not molecules. Therefore, we think that the use a polar NaF-NaOH extraction solution was not the reason for the observed results.

The tables and figures used to report the results are appropriate, though the figures are numerous.

While it is appropriate to report the 1:1 line in Fig7 (modeled vs. measured), I'm not sure why it is reported in Fig3. It seems to me the slope represents the proportion extractable. If I had to guess, the slope would be 0.58 (or its inverse). I'm not sure what it would mean to have the data fall on the 1:1 line in this case.

We agree that a visualization of the different efficiency of the two treatments in removing MOC does not need visualization with the 1:1 line. We will remove it.

I have become increasingly frustrated by the visual and qualitative interpretation of spectral data using a stack of "squiggly lines". While large differences might easily be apparent, smaller, more subtle differences, or large differences in smaller peaks may not be so

apparent. Differences among spectra should be tested quantitatively/statistically, perhaps using a multivariate method such as PCA or NMDS.

We agree the simple visual interpretation of the NMR-data from figures is largely qualitative. Therefore the figures were complemented by proper quantification of the relative contribution of C species (see Methods) to total OC in Tables 2 and 3. Since the overall result is that differences were small, scaling them into a PCA or NMDS would possibly lead to an over-interpretation of small differences probably not relevant for explaining differences in desorption or ^{14}C ages, and thus for the topic of this study. Therefore, we do not think that adding multivariate statistics would add further information to our conceptual understanding of MOC stability.