Response to review by Peter Kraal (Reviewer 4)

Original comments in black, responses in blue.

With interest I have read this manuscript, in which the authors explore the impact of adaptations to an established chemical method (reductive dissolution of Fe(III) by dithionite) to extract OC associated with Fe oxide minerals. This topic is of interest, because the impact of Fe-OC interactions have a bearing on the environmental fate (and possible global budgets) of both Fe and C. However, the manuscript seems to imply that no adaptation is necessary for the majority of marine sediment samples, and the full analytical impact of some suggested changes was in fact not explored. As such, in my honest opinion, I do not really see the added value of this "technical note". I do not think that the rather loose suggestion, that increasing dithionite concentration during OCFe extraction might be useful with the caveat that it may have negative consequences that were not investigated, is particularly useful to the geochemical community. Furthermore, I think that the authors overlook some key points that can be taken from the data and do not properly consider the relationships between the findings for the poorly ordered synthetic Fe(III) precipitates (and their properties) and natural samples.

We thank Dr Kraal for their review. In response to the overall value of the study, reviewer 3 raised a similar point which we responded to in detail in that review. Essentially, we have reframed the introduction and discussion around addressing the issue of comparability between extractions conducted by differing methods which we felt provided a better focus. We also think that within this context the synthesis we provide is more useful to colleagues utilising this methodology. The citrate issue is somewhat of a red herring and we detail why any potential implications of increasing citrate are likely to be unproblematic in response to Tom Jilbert, we also include an anoxic alternative to citrates increases thanks to suggestions by Susann Henkel. We are grateful for the comments around how we can better discuss the data we present; this was particularly useful for the discussion on freeze drying of the Arctic sediments which we have adapted substantially.

Below are detailed comments. Note: I prepared this review and afterwards read the excellent, extensive reviews by Susann Henkel and Tom Jilbert. I apologize for a limited degree of overlap.

Key points General language: unnecessarily verbose and at times rather vague, essentials are buried in winding sentences from which the reader has to deduce the actual information. For a technical note, the experimental section is poor. There seem to be some errors in the use of % and wt.% which are a bit confusing, please carefully check units for Fe and OC concentrations and extraction efficiencies.

This has been corrected following previous reviews; the use of "content" and "concentration" has been modified, and wt% has been used much more widely to avoid confusion.

Regarding the choice of organic compounds, I understand why the selection was made to have compounds with different amounts of carboxylic functional groups. It would be good if the authors could also explain the choice for these compounds in general, from a point of view of representing natural organic material in marine sediments. As mentioned by Tom Jilbert, the discussion on the impact of type of organic compound (L358 ff) is weak and too dependent on other study by Fisher et al.

The compounds weren't chosen to represent marine OM, but to be the simplest types of compounds capable of bonding with reactive iron. We have expanded on the selection of these compounds in the methods (line 110 in the original ms) through explicit mention of the bonding mechanism.

"These acids differ in their carboxyl group content (pentanoic- 1 COOH, hexanedioic- 2 COOH, 1,2,4-Butanetricarboxylic- 3 COOH), a factor thought to influence their binding to Fe$_R$ ***via bonding between carboxyl groups and mineral hydroxyls*** *(Karlsson and Persson, 2010, 2012;Mikutta, 2011)."*

Additionally, as Tom Jilbert also commented on this matter, for the discussion in line 358, we added the extra section (below) to expand on the role of OC compounds here and remove the reliance on the previous publication.

"Trends between Fe extraction and carboxyl content have been previously identified, with an increase in the number of carboxyl groups in an iron bound organic compound resulting in an increased proportion of Fe liberated from the sample (Fisher et al., 2020). This was explained by the greater amorphicity of ferrihydrite when carboxyl rich OC is incorporated into the mineral lattice, the resultant phase is less crystalline than pure ferrihydrite and therefore easier to reduce."

Coprecipitation is known to affect the structure of Fe(III) precipitates, was any mineralogical characterization of the Fe(III) precipitate performed? The impact of coprecipitation will perhaps be minor for 2-line ferrihydrite, i.e. a poorly ordered ordered Fe(III) precipitate would likely become a bit more disordered. However, there are that such minor changes in structure can result in large changes in reactivity (and thus solubility). And the very high OC:Fe ratio might in fact result in a Fe(III)/OC coprecipitated that is not 2L ferrihydrite but some organic-rich, amorphous hydrous ferric oxide. There is no mention or discussion on the likely characteristics of the Fe(III) precipitates and their relation to natural counterparts anywhere in the paper. Also, I wonder why only 2-line ferrihydrite was used? This poorly ordered Fe(III) precipitate often transforms very rapidly into more crystalline Fe(III) precipitates such as lepidocrocite or goethite (for which limited solubility in CBD would be much more relevant?). Overall, it would be good if the authors could spend some words on their exclusive choice for 2-line ferrihydrite (or the Fe(III) precipitate that would actually form under the experimental conditions): is it representative for the Fe-OC pool in soils and/or sediments?

Mineralogical characterisation was performed on these same samples in the previous study; we include the XRD analysis shown in sup fig. 1 of that paper below. The reviewer is correct that the mineral becomes more disordered and this is now mentioned in text (see response to previous comment). This XRD analysis shows that the resultant sample remains as ferrihydrite but the peaks soften with the decreasing crystallinity of the structure. Following the review by Susann Henkel, where the suggestion that the ferrihydrite may be transformed during freeze-drying was raised, we added a reference to this XRD data in text (end of line 334 in the original ms).

"An alternate hypothesis to describe the reduced Fe recovery for dried sediment, that transformation of ferrihdyrite occurred during freeze drying, was ruled out by x-ray diffraction (XRD) characterisation of the freeze dried phase as 2-line ferrihydrite (Fisher et al., 2020)."

We have added a section to expand on the use of 2-line over 6-line ferrihydrite. For the production of ferrihydrite we worked from the method of Schwertmann and Cornell (2000) *Iron Oxides in the Laboratory.* They describe the difference in production of 2-line vs 6-line ferrihydrite as

“Rapid, forced hydrolysis of Fe^{III} salt solutions under very acidic conditions at elevated temperatures (e.g. 80 “C) for a short period of time leads to 6-line ferrihydrite whereas rapid hydrolysis at RT and close to neutral pH produces 2-line ferrihydrite (Chukhrov et al., 1973; Schwertmann and Fischer, 1973)”. Additionally, 2-line ferrihydrite acts as a precursor for hematite and goethite, mirroring mineralogical transformations in sediment.

Regarding the choice of ferrihydrite, we know that in most natural environments the association of organic matter with Fe occurs during Fe^{2+} oxidation (e.g. Sodano et al., 2017, Riedel et al., 2013), and often (i.e., in many soils and most marine sediments) under low temperature, near netural pH conditions. These conditions are therefore much closer to the synthesis of 2-line than 6-line ferrihydrite, hence the choice. Additionally, 2-line ferrihydrite is used exclusively for OM coprecipitation as the high temperature, low pH conditions needed to precipitate 6-line ferrihydrite would make sorption of OM impossible due to hydrolysis. It would, without doubt, be interesting to observe what happens to the ferrihydrite-OC coprecipitate during further aging/crystallisation, but that is beyond the scope of this Technical Note.

In section 2.1 of the Methods, we added: *“2-line ferrihydrite was chosen as it is readily precipitated in the low temperature, circumneutral pH, aqueous conditions of marine surface sediments and has an established ability to be experimentally coprecipiated with organic matter (e.g. Eusterhues et al., 2008;Eusterhues et al., 2011;Eusterhues et al., 2014).”*

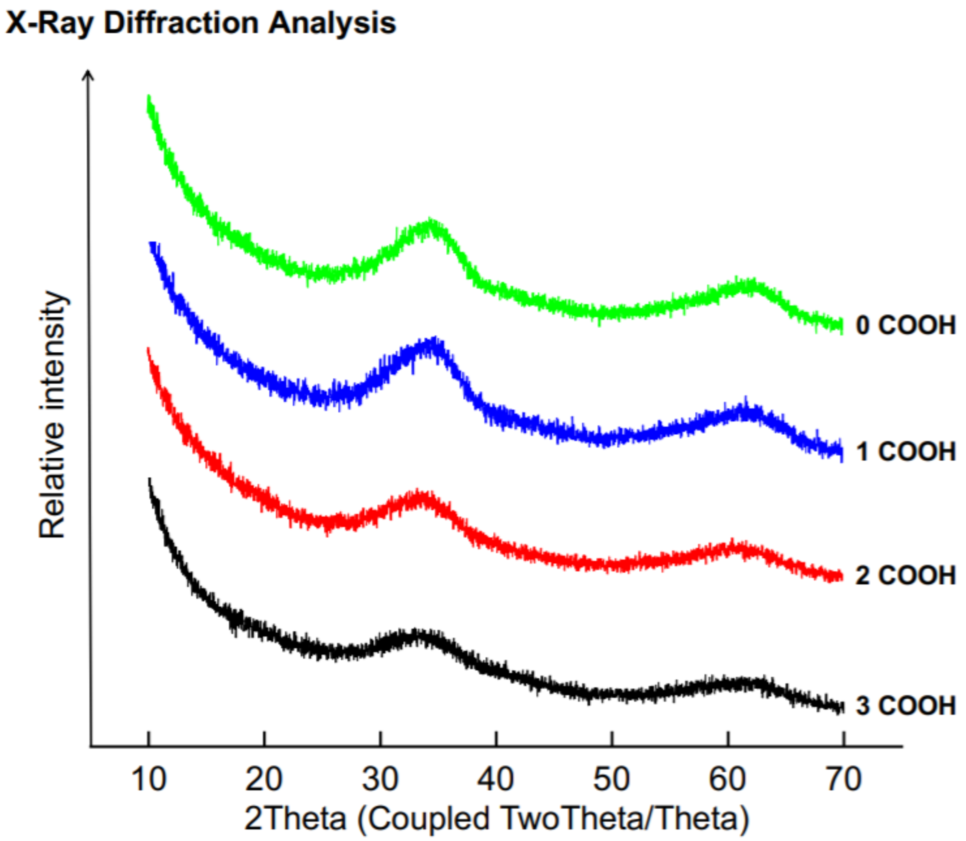

Figure S.1- Stacked XRD of coprecipitates with increasing carboxyl rich organic content.

Would have been interesting to see OC extraction efficiency for all treatments with the Fe-OC/sediment mixtures, not sure whether results from the variable dithionite experiment justifies assuming 1:1 relationship between Fe and associated OC extractability across the experiments conducted in this study. In fact, it would have also been very useful to have OC data for the dithionite/FeOC concentration range in the first experiment shown in Fig. 1: currently, we are not really given much to go by to understand how OC extractability varies as function of extraction conditions: one number for OC and then the assumption that Fe extractability is a perfect proxy for Fe-associated OC extractability for synthetic precipitates and natural samples alike. I find this a bit meagre (particularly for a technical note).

Unfortunately, it was not possible to produce OC data for all the points. The take away from Fig.1 is that even when Fe extraction is stretched to its maximum (by alteration of dithionite content), OC recovery is still incomplete. Therefore, at any point below this dithionite concentration, OC is certainly also incompletely extracted. While we do not state that Fe is a perfect proxy for OC loss, a strong similarity between the two values is to be expected given the 0.7-1:1 molar ratio between OC and Fe.

The impact of freeze-drying (section 4.2) is not represented properly. It shows a strong negative effect on the extractability of synthetic fresh Fe(III) precipitate, which is to be expected and reported regularly for poorly ordered Fe(III) precipitates (e.g. Kraal et al., Chemosphere, 2019). But is showed to have no effect on Fe extractability from the Arctic sample! However, in the discussion this important observations is ignored and there is a winding, unfocused and partly incorrect (see also Susann Henkel's review) discussion on the practices and challenges of freeze-drying.

We added to this section following Susann Henkel's review to include the suggestion of anoxic conditions. The inclusion of the Arctic sample is not meant for direct comparison with the artificial precipitate, but shows the effects of different dithionite additions on a natural sample representing what is often available to these types of studies. . It is also worth noting that our synthetic system is not meant as a direct analogy for real marine sediments, but plenty of studies perform mineral synthesis and utilise these samples for extractions, so even in synthetic samples this is an important difference. We did expand on the practicalities of marine sediment treatment in new sections added to 4.2

e.g.

"Wet thawed samples have been used more widely in the sequential extraction of Fe (e.g. Laufer et al., 2020;Riedinger et al., 2017;Wehrmann et al., 2014); additionally the Arctic sediment sample used in our analysis was similarly subject to freeze-thawing. However, our freeze-thawed sample showed no difference in its recovery for Fe compared to the dried variant of this sample indicating a potential interference from the thawing stage."

"However, the use of wet sediments is likely to be inappropriate for some analyses or sample sites either due to practical considerations such as the difficulty in transporting heavy wet sediments or when there is a need to preserve the sediment profile, for example protecting anoxic sediments from oxic redox transformations."

I have some reservations about the discussion on the impact of dithionite concentration on extraction efficiency in relation to increasing concentrations of Fe and OC. Firstly (L273-281), the authors present data from sediment/synthetic Fe-OC mixtures with Fe and OC concentrations that are incredibly high and as far as I know definitely not, as the authors claim, common (20-30 wt% OC and > 10 wt% Fe are not representative of normal marine sediments, shallow or deep).

We do not claim that we are trying to replicate normal marine sediment conditions here as this is a mechanistic study, see response to reviewer 3 for our discussion on high Fe contents.

Secondly, the authors focus a lot of attention on the relationship between dithionite concentration and OC content (L272-281), while I would expect that it is the ratio between dithionite and Fe (which is discussed later in the discussion, L282-291). I would argue that the bulk OC content of a sediment has relatively little to do with extractability of OC-FeHR, and I think this discussion on the role of dithionite concentration should better reflect the processes by which OC-FeHR is liberated. It actually seems like the "standard" (Lalonde) procedure works well for most sediments.

We agree that the ratio between dithionite and Fe is the most important fact; we do not discuss bulk sedimentary Fe since our samples contain only OC-Fe$_R$. The Lalonde method does likely work well for most sediments so we removed the direct criticism here, instead saying *"This finding demonstrates for these OC-Fe$_R$ rich sediments an increase in the dithionite content would aid a more accurate content determination."* However, many alternate methods have differing dithionite contents, and we show dithionite content can be an important factor Fe extraction.

The only adaptation really put forward in the current technical note is increasing the dithionite concentration, with two huge caveats: it seems only necessary for extremely Fe-rich samples (an observation not properly represented, as mentioned above) and the authors suggest that jacking up the dithionite concentration may have negative consequences on the performance of the extraction method, without actually exploring those potential negative impacts. I then wonder what to take away from this technical note?

As expanded on in the response to prior comments, the discussion around the increase of dithionite has been expanded to include a non-damaging anoxic method. The overall take away from this paper should be that methods which have already used differing dithionite conditions in the past are unlikely to be comparable in their results. That increasing extraction time, as done by recent methods, is largely pointless and that the important differences noted in drying of sediment is potentially important and requires further consideration. We hope this will become more apparent by the reframing of the discussion.

Detailed

L29. CO2

Changed

L87. HCl

Changed

L100-115. Would be nice to report the Fe/OC ratios during coprecipitation and the rationale behind choosing the concentrations. Also, how long was the precipitation allowed to occur?

Added: *"The resultant slurry was rinsed 5 times in 5 L of DI water **over 4 days**"*

Added: *"The mass of organic acids used was determined through batch coprecipitations with varying organic contents, the values used here represent the saturation point where a greater addition of organic does not result in increased association of the organic with ferrihydrite."*

Added: *“The coprecipitations produced three complexes with an increasing number of carboxyl groups resulting in an increased molar C/Fe ratio of (0.04, 0.25 and 0.70:1).”*

L111-112. a factor

Changed

L113. binding association? Seems repetitive. It’s also not type of binding, as all organic compounds have carboxyl groups. So. . . do you mean denticity, i.e. number of groups with which ligand binds to atom?

Changed to “strength of OC-Fe_R association”.

L112-115. Please rephrase this, it is unclear (what is “weak” binding in this context?) and the link to slurry and dry sample is not explained.

Changed to: *“For example, if it were shown that the OC-Fe_R containing was removable for a greater proportion of its Fe content in a slurried sate than as a freeze dried solid, is this related to the strength of the OC-Fe_R association or the physical difference (e.g. aggregation, surface area) of the dried vs slurried samples.“*

L117. “to explore whether mechanistic trends persisted”, please explain what this means here. In other words, use less fancy words and provide more concrete information about what you want to test by varying Fe-OC content

Changed to: *“The OC-Fe_R content of synthetic samples was varied to explore whether mechanistic any trends in Fe extractability for each of our experiments persisted at environmentally relevant OC-Fe_R contents.”*

The actual parameters we changed are detailed in their own methods section (2.4)

L119. “using the same original carrier sample and similarly treated to liberate OC and inorganic carbon”. You mixed marine sediment with your Fe-OC precipitate (I think that ‘carrier’ terminology is both unnecessary and incorrect as a carrier is strictly something else than a matrix). What does “the same original carrier sample” even mean, you used the same sample as used by others before? And can you please better explain what treatments were applied to the sample, rather than “similarly treated to liberate OC and inorganic carbon”, which could mean treatment for IC and OC was similar, or it was similar to treatments by Fisher et al. You are not writing a novel, the reader expects a clear and concise (and correct) description of materials and methods applied to them. So, as matrix (I am not adopting the ‘carrier’ terminology) you used sediment from which $CaCO_3$ and OC were removed, right?

Deleted “carrier”

We expanded this section following similar comments from an earlier reviewer to read:

“The OC-Fe_R content of synthetic samples was varied to explore whether any trends in Fe extractability for each of our experiments persisted at environmentally relevant OC-Fe_R contents. To achieve this we spiked the precipitate into a marine sediment sample from the Barents Sea (water depth 141 m; sediment core depth, 33.5 cm; station B6, E40; cruise JR16006). This sediment was freeze-dried then ashed (650 °C, 12 hrs) to remove OC and fumigated with HCl vapour to remove inorganic carbon.”

L121. Why “spiking” here and “mixing” before?

Mixing removed, see above.

L136-137. Was the artificial seawater deoxygenated? Do you expect any Fe precipitation issues when introducing oxygenated seawater into a sample with Fe(II)? Also, what was the composition/recipe of the ASW? Wash steps commonly involve simple strong salt solutions (e.g. 1 M $MgCl_2$), preparing ASW for this seems like an extra step for which I would like to know the rationale.

The artificial seawater was made with 35g/L of NaCl. It was not deoxygenated; Fe precipitation was avoided through acidification of the supernatants. The rationale for this is to remove residual citrate and bicarbonate as detailed in the Lalonde et al., (2012) method.

L140. "Testing the impact of different extraction conditions"

Changed subheading.

L162. What is freeze thawing? Just thawing?

Changed to "half was thawed following prior freezing at the point of sampling"

L174-184. Only after reading section 3.1 does it become clear in which samples OC was measured. Please rewrite/rephrase this section, the whole thing is difficult to understand. In particular, sentences like "Carbon content was not measured for all samples, but was used [used?] during the experiment where Na dithionite concentrations varied (see section 3.1)." and "This was performed to ensure that at the end point samples with incomplete Fe recovery also experienced incomplete OC recovery, as expected due to the 10%, at lower extraction efficiencies, please address this.

Rephrased the first part of this paragraph to make it easier to understand. The two quoted sentences have been removed.

"Carbon content was measured for the synthetic samples used in the experiment investigating the influence of Na dithionite content on Fe extractability (section 3.1). Here, C content was determined for all OC-Fe$_R$ contents (20-50%) both before and after Fe extraction to determine whether OC-Fe$_R$ was completely recovered; given that the Fe this OC is bound to was incompletely reduced across the series."

L206. At which dithionite concentration?

Added: "at the maximum dithionite content each sample was subjected to."

L208. "varying C content" is highly confusing here, you mean three types of OC, right?

Changed to "varying by organic acid"

L207-212. I am confused again. From the methods (section 2.1), I gather that the FeOC coprecipitates used in the dithionite concentration experiment reported in Figure 1 contained hexanedioic acid (' 2 COOH') and was performed with freeze-dried FeOC. In Figure 1, it shows a decrease in % Fe (and presumably % OC) extracted with increasing Fe-OC wt% from 90% to 40% between 20wt% to 50 wt% Fe-OC. But in Fig. 2, Fe extraction % decreases from 50% to 30% for 20 wt% to 40 wt% Fe-OC for freeze dried Fe-OC with 2 COOH as C compound. Am I misunderstanding something, or was there a big difference in Fe extraction % for the two experiments under supposedly similar conditions?

This is broadly correct, they are similar precipitates in that they were made in the same way, however, they were the product of separate precipitations (as one precipitation does not give a large enough yield). This results in variable Fe and C contents in the product of each precipitate, and the variation becomes stark when percentages are implicated (particularly at lower contents such as 20 wt%) due to the small numbers produced. So a relatively small absolute change can lead to a large % change. This illustrates why independent repeats would not be useful, information should be taken from the trends within each figure rather than a cross comparison between two different products. As noted, the relative trend (decrease in Fe recovery with increase in OC-Fe_R content) is the same, even if the scale differs.

As an aside, it would be very useful if the authors could mention at which dithionite concentration the second experiment was conducted (maybe it is tucked away somewhere in the Materials and Methods, please repeat in this section).

Yes, everything outside of the varying dithionite concentrations experiment was conducted according to the Lalonde method so the dithionite content returns to 0.25 g. Added *"Recovery of Fe following extraction with 0.25 g of Na dithionite"* to methods section 3.2

L215. With 2 out of 5 treatments not showing this trend, I would consider removing "typically".

Removed in earlier review.

L219-220. What is implied here? The sediment contains 20 wt% Fe-OC, that is not a trace amount where analytical limitations would interfere with trends (it is mentioned earlier that this is equivalent to 7 wt% easily reducible Fe). Please explain how the comparatively low (emphasis on comparatively, as the whole range is strongly biased towards really high Fe-OC concentrations) Fe content could obscure a trend.

It is correct to say these are not trace amounts, the issue with low concentrations in in the quantification of Fe. At 20 wt% OC-Fe, the Fe supernatant requires a 400x dilution to measure Fe in the 1-10 ppm range. Therefore, any small difference in the absolute amount of Fe measured would be multiplied by a large factor which results in the bias shown. This serves as an additional explanation for the difference in % terms referred to in the earlier comment.

L224. "were extractable for"? Please rephrase.

"Extraction of Fe from freeze-dried samples was…"

L225-227. No doubt this will be dealt with in the discussion (having read the discussion, I now know it is not addressed), but there is a large difference of the impact of freezedrying on freshly precipitated poorly ordered Fe (for which it is established that freezedrying decreases reactivity, likely by aggregation) and natural sediment from a depth of 22-23 cm in the sediment in which you will not find any labile, freshly formed Fe unless the sediment forms fresh minerals as an artifact of sample treatment (or am I wrong? There is in fact no information on the chemistry of the studied sediment at all). In this sense, there seems to be a mismatch between the synthetic sample and the natural test material.

We are soon to publish the chemistry of the sediment used here *(Faust et al., Nat. Comms, Accepted)*, hence why we did not include the analyses within this manuscript. A truncated sequential extraction was performed on this core up to 20 cm, at which depth we can confirm a low presence of iron oxyhydroxides (<0.15%), although this still represents 23.44% of the Fe_R pool, similar to the ~22% value we obtained. We have added this reference and a summary of sediment chemistry to the methods.

“This core has been analysed to a depth of 20 cm (see Faust et al., xxxx), where a reactive Fe content of 0.64 wt% was determined, of which 23.44% could be extracted by chemical treatments targeting labile and poorly crystalline Fe oxyhydroxides.”

We appreciate the overall point about the difference in Fe content of deeper sediment samples compared to surface sediments and have adjusted wording to remove any idea that the sediment sample is comparable to the synthetic minerals.

L230. “labile” sediment?

Removed

L231. Extraction time was extended in 15 min increments

Changed

L233. remains constant

Changed

L239. So there is the issue of OC hydrolysis at low pH. But to quantify OC in sediments, they are commonly decalcified to remove inorganic C, using dilute HCl. What do the authors think about this? Does low pH have to be avoided at all costs during CBD extraction, to then submit the sediment to low pH during decalcification?

This is a good point that we agree with, for that reason in this paper we chose to decalcify the samples by HCl fumigation as oppose to HCl rinses (as done in other studies). Since nothing is discarded from the sample when it is treated by fumigation then we reduce any possibility of OC loss.

L239. You targeted physical aspects? Such as. . . the concentration of dithionite??? This is chemical, surely.

Changed

L244-246. As mentioned before, the authors should probably also consider the large difference between the impact of freeze-drying on freshly coprecipitated Fe/OC and the impact of freeze-drying on relatively old and stable sediment. I do not think these are comparable.

See response to the final comment on the same point.

L258. Is the unit wt%? The context implies that this is the recovery efficiency (% of added Fe that is subsequently extracted), not the Fe content in the sample. (Should be checked in other instances as well, for instance y-axis of Fig. 1)

Checked and changed, Y axis of Fig. 1 doesn’t include wt%.

L268. “is removable for”?

Changed to liberates

L271-281. Because Fe and OC co-vary in the treatments, it is hard to judge what factor determines decreased extractability of OC-FeHR: is it the increase in OC or the increase in Fe (up to 24 wt% Fe!)? Merely looking at the trend in OC-Fe extractability in Fig. 1 does not answer that question. I would expect OC-FeHR to be liberated by dithionite because the Fe is reductively dissolved, and so the efficiency of OC liberation scales with efficiency of Fe reduction and not necessarily OC content. This would in fact mean that the bulk OC content of the sample is irrelevant, and the focus is on the wrong parameters in this section. The text in L293-306 supports this, the authors need to rethink their focus and wording here to better capture the chemical processes that occur when treating a sample with dithionite to reductively dissolve FeOx and associated compounds/elements. Again, regarding the units: the concentrations should probably be wt.% rather than % for the reported OC-Fe contents?

We are unclear exactly what this comment is getting at, we suspect this is an issue around phrasing which we hope to have resolved following earlier comments. We fully agree that Fe extraction is the critical parameter here, however in our experiments Fe is intimately linked to OC following coprecipitation. The bulk OC of the samples is the same as the amount of OC bound to Fe_R, which the comment agrees "scales with efficiency of Fe reduction".

wt% has been now used throughout to refer to OC-Fe_R contents.

Also, the authors mention that "many samples exist in the 20-30 [wt.]% range". And "the average value for marine sediment OC-FeHR composition is greater than 20 [wt.]%". These numbers and statements surprise me. Studying coastal and deep-sea sediments myself, I usually find TOC concentrations of 1-10 wt.% in a range of marine environments. The authors seem to claim that extremely organic-rich sediments are in fact very common (even though the phrasing "in many samples" is very different from "in many marine systems". . .), and then provide just one example, one value, from the Equatorial Pacific. I would like to have this point discussed a bit more: what kind of depositional environments host these very OC-rich samples, are we talking about modern or ancient (black shales and such, where the issue again appears that testing with labile OC-rich HFO makes little sense)?

The samples we are referring to here as having over 20% OC-Fe_R contents mean that 20% of the TOC pool of a sediment is bound to reactive iron, not that 20 wt% of the sediment is OC-Fe_R.

L293-306. So, now the authors are saying that the existing method works fine for most samples, that increasing the concentration of CBD could work for some (extreme) samples but may also have negative effects that are not explored. . . So what are we to do with this technical note?

This narrative can be shifted by changing the focus of the paper slightly. We can instead say that the wide range of dithionite concentrations used in previous methods (Table 1) can have an effect on Fe extractability, and therefore studies with different dithionite masses can produce non-reproducible results, rather than focusing on achieving maximal extraction. The point about citrate as a negative was addressed in response to Susann Henkel and Tom Jilbert, essentially we now include the prospect of anoxic conditions for high Fe samples (which Henkel et al., (2016) shows to work). The issue with citrate contamination is also of minor importance for natural samples, it is only synthetic ferrihydrite which has a much greater free surface (less OM coverage) where citrate contamination can become a big problem. Lalonde et al., (2012) showed that citrate contamination for natural sediments is at most 0.08% of dry sediment weight so increasing it is unlikely to significantly change this. We have toned down the caution given to citrate in the manuscript and included the anoxic alternative.

L307-367. This is an excessively lengthy paragraph on freeze-drying. It is actually strongly detached from the findings and just meanders along various aspects of freezedrying. The main issue is that the authors do not correctly represent their own results: a decrease in Fe extractability was found in the synthetic samples, but not in the Arctic samples! As I mentioned before, it is to be expected that freeze-drying a much stronger effect on a fresh, poorly ordered Fe(III) precipitate than on a rather old sediment sample. In fact, there are findings that show that freeze-thaw cycles can increase extractability of elements (the authors also touch upon this, and Susann Henkel also hints at some inconsistencies in this section). Note: this is a change in extractability, not content; the text became rather confusing when the authors started to speak about freeze-drying as a treatment that can increase the contents of for instance OC and metals. . . (L351-353). Overall, this whole section fails to address the key point, i.e. the discrepancy between the results for the synthetic and environmental samples (or, in broader terms, the difference between artefacts in fresh and old Fe minerals, whether they are synthetic or natural), and instead presents long and rather unfocused and at times confusing literature review on freeze-drying.

Despite my sharp tone (work in progress), I trust this review is fair and constructive.

We have rewritten this part of the manuscript to remove the extended discussion on freeze-drying. We state that the dominant Fe phases here are crystalline phases unlikely to be extracted by CBD treatment under any circumstance. We also consider the fact that in deeper sediments such as ours the freeze drying effect we observe for freshly precipitated ferrihydrite has occurred through natural aging of the sediment (due to transformation and aggregation with diagenesis). This allows us to conclude that extraction of surface sediments with high ferrihydrite contents is likely to be improved through use of non-freeze dried sediments while for those deeper in the sediment profile these effects are less important.

“Wet thawed samples have been used more widely in the sequential extraction of Fe (e.g. Laufer et al., 2020;Riedinger et al., 2017;Wehrmann et al., 2014), additionally the Arctic sediment sample used in our analysis was similarly thawed following freezing on collection. However, our thawed sample showed no difference in its recovery for Fe compared to the dried variant of this sample. Natural aging processes within the sediment could describe this lack of freeze-drying effect in deeper sediments (here 22-23cm) with the transformation of labile poorly crystalline ferrihydrite to more stable phases which have increased resistance to CBD extraction. The loss of bound OM through biological degradation or loss in mineral transformation (Jelavić et al., 2020), as well as the decrease in redox potential at depth effect mineral stability beyond what is represented by synthetic precipitation of fresh ferrihydrite complexes. ”

“We conclude that freeze-drying of fresh or synthetic sediments containing poorly crystalline iron is likely to negatively bias the quantification of these phases by chemical extractions. This may be of particular interest where synthetic iron minerals are used to calibrate extraction protocols. In deeper sediments where the iron content is dominated by more crystalline phases, the effect of sample preparation is no longer significant. We expect that the scale of the freeze-drying effect decreases in line with the decreasing oxyhydroxide content of the uppermost part of the sediment profile, however, it may be useful to determine the threshold at which sample preparation no longer imparts a quantification bias on Fe in future work.”

We have removed the section relating to changes in absolute C/metal contents. We hope the changes we have made in this section satisfy the request for a clearer discussion on the differences between synthetic/environmental samples.

Additional references:

Sodano, M., Lerda, C., Nisticò, R., Martin, M., Magnacca, G., Celi, L. and Said-Pullicino, D., 2017. Dissolved organic carbon retention by coprecipitation during the oxidation of ferrous iron. *Geoderma*, 307, pp.19-29.

Riedel, T., Zak, D., Biester, H. and Dittmar, T., 2013. Iron traps terrestrially derived dissolved organic matter at redox interfaces. *Proceedings of the National Academy of Sciences*, 110(25), pp.10101-10105.

Faust, J.C., Tessin, A., Fisher, B.J., Zindorf, M., Papadaki, S., Hendry, K.R., Doyle, K.A., März, C. Accepted. Millennial scale persistence of organic carbon bound to iron in Arctic marine sediments. Nature Communications.

Laufer, K., Michaud, A. B., Røy, H., and Jørgensen, B. B.: Reactivity of Iron Minerals in the Seabed Toward Microbial Reduction – A Comparison of Different Extraction Techniques, Geomicrobiology Journal, 37, 170-189,

Riedinger, N., Brunner, B., Krastel, S., Arnold, G. L., Wehrmann, L. M., Formolo, M. J., Beck, A., Bates, S. M., Henkel, S., Kasten, S., and Lyons, T. W.: Sulfur Cycling in an Iron Oxide-Dominated, Dynamic Marine Depositional System: The Argentine Continental Margin, Frontiers in Earth Science, 5

Wehrmann, L. M., Formolo, M. J., Owens, J. D., Raiswell, R., Ferdelman, T. G., Riedinger, N., and Lyons, T. W.: Iron and manganese speciation and cycling in glacially influenced high-latitude fjord sediments (West Spitsbergen, Svalbard): Evidence for a benthic recycling-transport mechanism, Geochimica et Cosmochimica Acta, 141, 628-655

Henkel, S., Kasten, S., Poulton, S. W., and Staubwasser, M.: Determination of the stable iron isotopic composition of sequentially leached iron phases in marine sediments, Chemical Geology, 421, 93-102

Jelavic, S., Mitchell, A.C., Sand, K.K. Fate of organic compounds during transformation of ferrihydrite in iron formations. Geochemical Perspectives Letters, 15, 25-29.