Point-by-point changes and responses to comments

Referee Timothy J. Heaton

Dear Timothy J. Heaton,

First of all, we would like to thank you very much for reviewing our manuscript for a second time. We appreciate your comments on improving the revised version of our manuscript.

We provide responses to each individual comment and our suggested changes and adjustments to the current manuscript below. We have highlighted your comments in black and *italics* and highlighted our responses in blue.

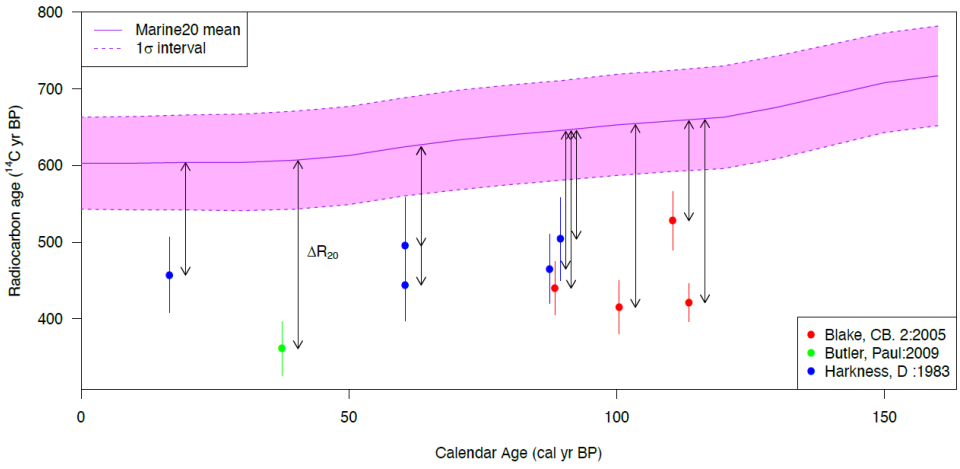
Thank you once again for taking the time to review our study.

On behalf of all the authors,

Gregor Pfalz

Reservoir Ages

I remain slightly confused by how they have worked out 14C reservoir ages. Calculation of the reservoir age would require them to have a 14C (marine/lake) sample (at the top of the core) for which they know the calendar age (e.g. 10 cal yr BP = AD 1940). The total reservoir age/hard-water effect is then the difference between the 14C age of this (AD1940) sample and the IntCal20 estimate at 10 cal yr BP (AD 1940). For a ΔR we instead look at the offset to Marine20, i.e. the difference between the 14C age of this sample and the Marine20 estimate at 10 cal yr BP. This is shown below. It is the vertical difference between 14C ages for a fixed calendar age:



Now for a lake core, if we don't have a measurement at the top of the core (which you know/assume has a calendar age corresponding to the core drilling) you can't do this. You need a sample of known calendar age.

My understanding (??) from their text is that what they suggest instead is to:
a) Initially calibrate all their 14C dates assuming no reservoir effect against e.g. IntCal20. This provides (non-reservoir) calendar age estimates for each sample.
b) Fit a model through the resultant (non-reservoir) calendar ages and estimate the calendar age of the top of the core
c) Estimate the reservoir age as the difference between the calendar age at the top of the core and the year the core was drilled.

Have I understood this correctly? If so, this is effectively working out the horizontal difference (in cal years) between the sample and the curve in the figure above. This will give very slightly different values (as the IntCal curve is wiggly and not quite y = x)).

***Importantly**, if you do not have a 14C sample at the top of the core of known age, then you cannot do the first approach. Their proposed approach is therefore probably the best one can achieve as an approximation. I am not therefore suggesting they do anything differently (especially as they add an uncertainty).*

***Suggested Action**: If my above understanding of their approach is correct, they might just wish to state this is an approximation (as it is all that it is possible), e.g., on line 164 just swap calculate for approximate (everything else can probably stay as it is)*

We approximated the reservoir effect by subtracting the target age from the mean predicted age, whereas the associated error we based on the two-sigma uncertainty ranges of the prediction

If my understanding is not correct then they do not need to do anything (it may just be the fact I have not quire understood).

> Thank you for your comment and suggestion. Yes, you understood our approach correctly. We agree with your suggestion and changed the sentence (Lines 163 to 164, Page 5) to: "We **approximated** the reservoir effect by subtracting the target age from the mean predicted age, whereas the associated error we based on the two-sigma uncertainty ranges of the prediction."

Case Study 2 (CS2) – I feel much more strongly here that it is dangerous to suggest users can fit age-depth models to data like this. In my opinion, a much clearer/stronger caveat is needed

Despite the author response, I still feel as though there is a significant danger that some readers will see case study CS2 and think LANDO will automatically and always solve issues of the divergent dates. I am still not sure that the LANDO ensemble covers all the potential options as it

seems to have decided that the 14C dates around 400-500cm are all erroneous (no models extend up to here). Further BChron produces very bizarre sedimentation rates.

I understand the authors wish to use it as an example to show how the methods can provide very different age-depth models on pathological examples. That is fine so long as it doesn't give the impression to LANDO users that they can treat it as a perfect black-box. I fear the manuscript still gives an idea, especially lines 564 – 570, that LANDO and proxy optimisation can automatically fix datasets where the calendar ages are in huge disagreement.

The authors have added a caveat on lines 555 – 560, however I believe something stronger is needed here, with a similar statement in subsection 3.2 introducing this example. Placing the caveat only in the discussion (on line 560) means many readers will miss it. This caveat should say that:
a) the first thing one should do in such an example of incongruent dates (and outliers) is to go back and use their expert knowledge to assess the reliability
b) use extreme caution fitting any age-depth model.
c) LANDO will not automatically fix the issue of divergent dates (even when the whole ensemble of 4/5 models is considered)

***Suggested Action 1**: Again, the authors need not redo anything but I would like to see something making the above points clear on pg17. For example, as the introductory paragraph:*

3.2. "Inconsistent sequence" – Case Study Number 2
For the second case study we consider an example where the underlying calendar age estimates within the core are highly incongruous with one another, see Figure 3. Before considering modelling the age-depth relationship using data that disagree so strongly, a user must explore and aim to understand the reasons for any outliers. Fitting any age-depth model, including the LANDO ensemble, to such divergent data should be done with extreme caution and we do not recommend it. Here we primarily aim to illustrate the range of age-depth models obtained within the ensemble, and the process of optimisation using proxy-based lithography.

***Suggested Action 2**: I would also remove the paragraph from lines 564 – 570 which could be read to suggest that the LANDO ensemble will always fix things. I do not see it is the case that a LANDO ensemble will include all possible age-depth relationships (both generally and for this specific example). One should really not be fitting automated models to this kind of data.*

I do not think adding such caveats will affect the popularity/use of LANDO, since hopefully datasets as incongruous as CS2 are rare. It will however hopefully prevent bad inference based on bad data.

> Thank you for the detailed comment and your concerns. As you rightly pointed out, we wanted to show CS2 as an example, but not as the ultimate solution to scattered data. We agree with your first suggested action and have added the following paragraph on page 17: *"For the second case study, we considered an example where the underlying age determination data within the core are very contradictory to each other (see Figure 3). Before considering modeling such an age-depth relationship with conflicting data, users*

need to investigate and try to understand the reasons for any outliers. Fitting any age-depth model, including the LANDO ensemble, to such divergent data should be done with extreme caution and we do not recommend doing so without further deliberate investigation. Here we primarily aim to illustrate the range of age-depth models obtained within the ensemble as well as the results of the optimization with our proxy-based lithology."

Regarding the second proposed action, we removed most of the paragraph, except the last sentence, which says:

"For palaeoenvironmental reconstruction, users should also propagate these increased uncertainties into their proxy interpretation, which is often underrepresented (Lacourse and Gajewski, 2020; McKay et al., 2021)."

We added this sentence to the previous paragraph, because with this statement we want to ensure that users account for age-depth model uncertainty in their further analysis, regardless of which models they use.

Line 40, Page 1: Change "Lakes" to "Lake (singular)

Changed.

Line 71, Page 2: Change "Implemented methods" to "Methods have been implemented"

Changed.

Line 72, Page 2: Add "to" before "support"

Added.

Line 80, Page 2: Change "usually, users only select" to "user usually only select"

Changed.

Line 149, Page 5, Table 1: Should this be "or" rather than "and" to correspond with age above?

Correct - we changed it to "or".

Line 174, Page 6: Change "create" to "creating"

Changed.

Lines 185 to 186, Page 6: Split sentence into two sentences to "… modeling outcome. This is challenging if no objective prior knowledge exists"

Agreed. We separated the sentence into two sentences: *"As mentioned before, the selection of model priors and parameters has an impact on the* ***modeling outcome. This is challenging if no objective prior knowledge exist.****"*

Lines 192 to 193, Page 6: Change to "…for problematic cores, as recommended by Blaauw et al. (2021)." so it agrees with author response

Changed.

Line 206, Page 7: Add "an" before "attribute"

Added.

Line 238, Page 8: Change commas to "As a counterexample, for the second study we have chosen"

We changed the sentence to: *"**As a counterexample, for the second case study we have chosen** the sediment core EN18208 (Vyse et al., 2020)."*

Line 293, Page 13: Change "extract" to "extracted"

Changed.

Line 535, Page 25: Add "a" before "significant"

Added.

Line 664, Page 28: Change to "provides neither a query optimizer, nor relies on Map-Shuffle-Reduce"

We changed the sentence to: *"The key difference is that "Dask" **provides neither a query optimizer, nor relies on Map-Shuffle-Reduce,** a data processing technique for distributed computing, but instead uses a generic task scheduling (cf. Dask Development Team, 2016)."*

Line 684, Page 29: Change "result" to "resulting"

Changed.

Line 696, Page 29: Change order to "can increase this factor even further by…"

We changed this sentence to: *"When considering that our setup consisted of six CPUs (12 threads) and 16 GB RAM, user **can increase this factor even further by** using larger computing facilities."*

Line 705, Page 29: Change "bares" to "bears"

Changed.

Line 713, Page 29: Remove "do not" and "more"

Removed.

Line 726, Page 30: Change “most” to “best”

Changed.

Line 733, Page 30: Change “most” to “better”

Changed.