Review of:
*Towards the construction of regional marine radiocarbon calibration curves: an unsupervised machine learning approach*

*Geochronology Discussions*
*Feb 2024*
*Referee: T J Heaton*

## Summary

This is an interesting, and stimulating, paper which aims to use model simulations to identify ocean regions which appear to have similar levels of $^{14}C$ depletion. The authors wish to make the case for regional marine radiocarbon calibration curves. They suggest that one might aim to identify suitable partitions for those regions using output from a computer model – inherently trying to cluster together locations for which the models generate similar output together.

I think this is an interesting and novel idea, although there are definitely challenges to overcome. Most directly, whether the current ocean circulation models are capable of providing reliable regional estimates and how/whether they account sufficiently for the degree of uncertainty regarding the behaviour of the oceans, fine-scaled circulation, the extent of sea-ice, and past changes to the carbon cycle.

In this paper the authors consider two ocean circulation models. One models a single year at high temporal precision and so I am not entirely sure if this is particularly relevant to a radiocarbon calibration curve providing at-best annual resolution estimates of oceanic depletion (but more likely somewhat coarser). The other model (UVIC) can be run over longer periods of time (they run it for about 15,000 years) and so would seem much more relevant to the problem at hand.

They run UVIC under two different scenarios with the aim of seeing if the regional clusterings of the set of simulated surface ocean depletion time series is consistent with different forcings. While running different carbon cycle scenarios certainly captures some potential variability, it is unclear whether other aspects of the model remain the same, and therefore to what extent there remains model specificity in the clusterings. I am not an expert in the details of the modelling, but I presume some of the parameters/interactions/fundamental underlying ocean circulation structures remains similar in both runs reliant even though they are inherently somewhat uncertain. As such one might need to be slightly cautious that other models might provide different clusterings.

Interestingly, both UVIC scenarios suggests similar clusters – effectively:

1) High-latitude (polar) Southern Ocean;
2) Pacific Ocean basin;
3) Atlantic Ocean basin;
4) High-latitude (polar) Arctic Ocean

These are the most immediate partitions one would select (indicating that the underlying UVIC model is doing sensible things).

However, there are perhaps some less expected features in that the regions in each cluster are not always geographically connected to one another (and in some cases even lie in different

hemispheres). I think this raises some practical questions as to whether such highly-disconnected locations should be grouped together into a single regional calibration curve (especially as the authors implicitly propose geophysical ideas as to why that general clustering, e.g. of the Southern Ocean, is appropriate in some of the discussion). Or whether we might need a combination of expert knowledge and ML to define regions.

I have no concerns with the application of the statistics. I think the paper will certainly provoke discussion and new ideas in the community. I certainly enjoyed reading it and it made me think. I would therefore recommend its publication.

I do have some comments, questions and suggestions that I hope will start discussion. I lay these out below. In general, my view is that regional marine calibration curves would be a fantastic and hugely important achievement, however we are still quite a way off being able to reliably generate them. This paper provides suggestions as to how we might navigate our way towards them.

---

**Main Points:**

*Existing/Related IntCal Recommendations on Regional Marine Calibration Curves:*

The IntCal group have aimed to explain to the community that using the current MarineXX curves for calibration requires the application of significant simplifications/approximations. In particular, users are generally required to estimate a value of $\Delta R_{20}(\theta)$, the offset between the localised surface-ocean depletion and the Marine20 curve, based on modern-day values and then consider that $\Delta R_{20}$ remains roughly constant over time. The idea here is that the main changes in oceanic $^{14}$C depletion occur at a global scale. This approximation of a constant $\Delta R_{20}(\theta)$ is discussed in detail in Heaton et al. (2023a) along with a broader discussion of the limitations of the Marine20 curve (which includes some of the issues the authors identify here).

Calibration of marine $^{14}$C samples is particularly challenging for polar (high-latitude) oceanic regions during glacial periods (as the authors highlight in Fig 1b) as modern day $\Delta R_{20}(\theta)$ are unlikely to be appropriate. During these times (and in these high-latitude locations) there may have been periods with substantial sea ice (that came and went) but which is not present in the modern day. This sea ice would have caused further localised $^{14}$C depletion and an increase in the MRA that is not represented in the global-scale Marine20 curve

To address this, the IntCal group have, in fact, already proposed a method to effectively perform regional marine calibration at high-latitudes. This can be found in Heaton et al (2023b). The proposed approach is a very simple (approximate) way of estimating upper and lower bounds on changes in $\Delta R_{20}(\theta)$ in glacial periods that uses the regional output of the LSG OGCM (Butzin et al. 2020).

At a given latitude, in Heaton *et al* (2023b) we propose first estimating a modern-day $\Delta R_{20}$. However, at high-latitudes during glacial periods, we recommend that this modern-day value may be too small and users will need to consider by how much it may be increased in a high-depletion $^{14}$C scenario. The amount by which we suggest it may need to be boosted is region (specifically latitude) dependent. We advise users with samples from glacial periods to calibrate under both a high (boosted- $\Delta R_{20}$ and low (modern day $\Delta R_{20}$) depletion scenario to provide bracketing calibrated ages which hopefully encapsulate the true calendar age. This latitude-dependent adjustment effectively matches the partitioning proposed in this ML paper (where the clusters are effectively latitudinal bands concentrated on polar regions).

This proposed bracketing approach is not however mentioned in the submitted manuscript, and I think the representation of current calibration approaches might suggest to a reader that no suggestions to overcome the challenges of Marine calibration exist. The proposed Heaton et al. (2023b) approach is certainly simple, and coarse, but provides a first option until we get to a point where detailed and reliable regional marine calibration curves are possible. It may not allow highly-precise calibration (due to the high-level of uncertainty on past sea-ice extent) but hopefully will provide accurate calendar dating in polar regions.

<u>*Notation:*</u>

I would suggest that it would be extremely useful to add a subscript on all your estimated values of depletion to denote which calibration curve these are measured against, e.g., $MRA_{20}(\theta)$ and $\Delta R_{20}(\theta)$ if you are measuring against the IntCal20/Marine20 set of products, $MRA_{13}(\theta)$ if you are using the IntCal13/Marine13 products. See Heaton et al. (2023a) for our IntCal group advice/suggestion on this.

Without a subscript, it is unclear whether the plotted estimates of MRAs relate to the most recent set of IntCal20 curves or previous ones (as, e.g., Menviel *et al.* 2015 used for Figure 1b, would have originally been comparing with IntCal13 rather than IntCal20). Of course, the true marine depletion/offset is independent of the calibration curves, but the estimated values are not. The estimates will change with each update of calibration curve

This is particularly important when considering changes in $\Delta R(\theta)$ over time, as in the glacial period the Marine20 curve has changed significantly from Marine13. This will greatly affect the respective evolution of $\Delta R_{13/20}(\theta)$. The atmospheric curves have also changed somewhat which will affect the overall MRA but to a lesser extent.

Specifically, in Figure 1b, if these plots relate to changes in $\Delta R_{13}(\theta)$, then they could be quite different now. I am assuming that the main differences in the value of $\Delta R_{13}(\theta)$ at a specific location (i.e. the values plotted) will be during the late glacial. Here Marine13 assumed a constant global-scale MRA whereas Marine20 does not.

If Figure 1b is plotting $\Delta R_{13}(\theta)$ (i.e., using Marine13) then it may be that the maximal variation in Equatorial regions is now much smaller with $\Delta R_{20}(\theta)$ and Marine20. It would also be interesting to see if the maximal variation in $\Delta R_{20}(\theta)$ at higher latitudes with UVIC are similar to the suggested boosts/shifts proposed in Heaton et al. (2023b) using the LSG OGCM.

<u>*General Philosophical Question:*</u>

A philosophical question I might pose with the proposed approach is whether, until we are sure that the computer models accurately represent ocean circulation and past carbon cycle changes, can we use them to reliably cluster the ocean into regions where we can reliably group observational $^{14}C$ data to create regional marine calibration curves?

On the other hand, if we are confident that these models can represent circulation and carbon cycle, do we actually need data or can we just use the models themselves to produce calibration curves for any chosen location?

I would assume that to create the regional curves we would ideally collate records from multiple locations in any cluster. If so, do we need the prior (model-determined) clustering, or could we just create a lot of location specific curves from that data? Is a side/main benefit of the identified clustering is that it can to tell us about underlying properties of the computer models rather than to generate calibration curves?

*What to cluster on?*

A key question seems to be whether it is critical to generate regional marine calibration curves that need no $\Delta R$ adjustment for any location within that region (i.e., all locations have the same level of $^{14}$C/depletion)? Or is it more critical to generate calibration curves which might all need adjustment, but that adjustment is constant over time (i.e., they get the right evolution but perhaps offset by a constant amount).

I am little unclear which of these two options is being aimed for? If the concern is in the non-constancy of $\Delta R(\theta)$ over time, should the clustering be done on the variability rather the absolute value of the MRA. Is this the distinction between the normalised and un-normalised approaches to clustering. Are all the simulated outputs (for all locations) set to have mean zero for the normalised approach to clustering (i.e., a constant offset is considered identical) or is there also some renormalising of variance at every calendar age?

*Sea Ice*

My expectation is that a substantial factor in determining $^{14}$C depletion in any marine location is sea ice. This would seem to be highly regional during glacial periods. Do we require detailed knowledge of sea ice extent and location for the computer models to reliably cluster locations together?

*Practicality of Disconnected Locations within a Cluster*

It seems potentially controversial to suggest one might create regional curves that cluster/group locations which aren't geographically close together (e.g., Fig 5 has clusters that contain high-latitude NH locations and high-latitude SH locations). Also, Fig 12 has some high-latitude NH regions which are clustered with Antarctic waters. It would be interesting to know whether this would be supported by the practical oceanographic community.

This is briefly discussed in Figure 14. Here, cluster 2 predominantly represents Antarctic waters but with a few regions in the high NH too. You implicitly seem to propose a potential link between the increase in MRA observed in this region to the Antarctic temperature. Are you suggesting this is perhaps the increased presence of sea ice? If it is sea ice in the SSoutehn Ocean then would affect the NH regions in the same cluster?

Is it appropriate to create a single regional curve for such geographically distinct regions? How might one strike a balance between a black-box ML clustering and incorporation of expert geoscientific knowledge?

---

**Technical Comments/Questions:**

**Clusters/Continuum** - How much of the variation is really on a continuum, and how much is it there really are distinct and separate clusters? The plots in Fig 5 suggest the clustering appears mainly based upon a scale of the mean overall MRA rather than hugely different shapes.

Interestingly, the clusters are predominantly latitude-based, is this basically indicating that the UVIC model has $\Delta R$ increasing more in polar regions than in more equatorial regions in glacial periods due to sea ice in those high-latitude locations? If so, this suggests UVIC and the LSG OGCM model concur with one another.

**Figure 14** – I do not entirely understand this plot. What is the difference between the red and blue lines for each identified cluster? You say they both represent the cluster – but isn't the point that there is supposed to be a single MRA for all sites in the cluster (not two). Also are the means (shown in solid lines) the UVIC model output or the averages of the observed data. I am presuming the latter, if so, are the observed MRAs in the specific clusters actually that similar – they seem to vary by 1000 $^{14}$C yrs between records within a cluster.

**Figure 13** – It is certainly the case that correct clustering /partitioning will provide you with more precise calibration curves. However, this seems a rather unfair comparison of Marine20 against the clustering approaches to make that point. In the central and RH panels, it seems you are effectively comparing simulations from UVIC with themselves; whereas in the LH panel you are comparing UVIC simulations with another entirely different model BICYCLE/Marine20. This is never likely to do as well. Furthermore, Marine20 aims to incorporate a much wider range of climate scenarios than the single climate scenario represented in the other panels by U-Tr.

**Specific (Minor) Comments:**

***Line 42*** – I would say that perhaps the NH atmosphere is the only reservoir for which we have an entirely robust curve based upon direct observation (and even this is somewhat reliant upon the DCF in Hulu Cave being constant over time once we go back further than 14,000 cal yrs).

The SH calibration curve is, in large parts, reliant upon NH data and an assumption that the interhemispheric $^{14}$C gradient (IHG) has been roughly constant over time. Of course, the IHG is expected to be much less variable than the $^{14}$C depletion in the surface oceans, but we do still need more SH reference material to increase the precision of SH calibration.

Suggest one could reword the intro slightly to make clear that the SH calibration curve is certainly still a work in progress and more reference data is needed (and in fact, even the NH curve is reliant upon quite strong assumptions)

***Figure 1*** – Panel a: Suggest you could be much clearer about precisely when this is a plot of in the title of the plot and the caption (also in the text you say it is pre-industrial, whereas in the caption you say it is modern and bomb-corrected? Which is it? Can you give a specific date as the overall MRA is highly variable from one year to the next. Also, it is unclear if this is modelled output or observation based – suggest could clarify what GLODAP is? Panel b needs clarification if this plots the changes in the estimated $\Delta R_{13}(\theta)$ or $\Delta R_{20}(\theta)$ (as explained in main comment).

***Line 261*** **–** Do you mean cluster 1 forms a latitudinal band? Not longitudinal band

***Useage of the Term "Data"*** **–** In general, I feel it would be useful to distinguish through the manuscript between genuine observed data and model output/simulations. Personally, I would restrict the use of the term *data* to refer to when one has actual observations. I would not describe output from a model as data – I think it is better to refer to it as modelled output, or a time series vector of simulated values. For example, I would not say you are clustering data (as that may suggest to readers that there are underlying observations) but rather you are clustering the vectorised model output.

***Figure 2*** – Suggest it could be made clearer this is an entirely artificial example to illustrate what clustering is. Perhaps this could be achieved just by creating a subsection explicitly called "A simple illustration of clustering" into which it could go. Initially I was a bit confused if these were the clustering of the actual vectorised simulated time series (with the principal components as the two plotted axes). Also, it would seem for Fig 2 as though 3 clusters is most appropriate to represent the data, rather than 4, so a bit unclear how it fits with the surrounding section about how you chose the optimal number of clusters.

---

**References:**

Heaton TJ *et al.* (2023a) 'A Response to Community Questions on the Marine20 Radiocarbon Age Calibration Curve: Marine Reservoir Ages and the Calibration of $^{14}C$ Samples from the Oceans', *Radiocarbon*, 65(1), 247–273. doi:10.1017/RDC.2022.66.

Heaton TJ *et al.* (2023b) 'Marine Radiocarbon Calibration in Polar Regions: A Simple Approximate Approach using Marine20', *Radiocarbon*, 65(4), 848–875. doi:10.1017/RDC.2023.42.

Butzin, M., Heaton, T. J., Köhler, P., & Lohmann, G. (2020). A Short Note on Marine Reservoir Age Simulations Used in IntCal20. *Radiocarbon*, *62*(4), 865–871. https://doi.org/DOI: 10.1017/RDC.2020.9