We thank Reviewer 1 for the constructive and thorough comments, especially regarding the implementation of the open-source software and the detailed software review. A point-by-point response to each comment is detailed below (blue).

This paper presents an interesting evolution of the the previous d18O alignment and stacking functions. The main advancements are the the use of an empirically derived sedimentation rate prior and incorporation additional age information including radiocarbon dates and tephras, tie points, etc. While I think this will be a fine contribution to the field, I think the manuscript needs to be further clarified. The bulk of the background focuses on d18O alignment. The radiocarbon descriptions can be somewhat abbreviated and overly simplistic.

The main issue I have is that the information from which the prior was obtained is completely absent. The reader knows nothing about the sediment cores, their locations, age ranges, and depositional environments. An empirically derived prior that replaces tunable parameters is only an advancement if it is appropriate to the readers sediment core. Statements around Line 220 seem to indicate that the prior is a poor match for the data, and the prior, not the radiocarbon dates are the greatest influence on the age model in the radiocarbon-dated interval. Hence, I recommend the discussion of the prior be greatly enhanced.

Thank you for bringing this to our attention. The prior presented here is derived from the same compilation of cores as Lin et al., (2014), i.e., HMM-Match. We recognize the importance of transparency, and in the revisions, we will include a thorough description of these cores, criteria for inclusion, and the methods of construction. In addition, we will add a section in the discussion that describes the potential strengths and weaknesses, the appropriateness of this data to be used as a prior for any ocean sediment core, and the option of users to substitute alternate priors. Also, our statement on line 220 was unclear and has been misinterpreted; see below.

I suggest the authors take a close look at spellings of acronyms and names. I noticed several different (and incorrect) spellings of Obrochta, as well as some acronyms that were transposed. I've noted the former in the below line-by-line comments.

Thank you, we will fix this.

Finally, the manuscript needs to state the system requirements for running the software. I note that it uses a parallel for loop, which is not included in the standard matlab distribution. So users without the Parallel Computing Toolbox cannot use this. Also of course the sampling of PDFs will require the stats toolbox. In my experience, most people have the stats toolbox but fewer can run a parfor loop.

While the system requirements are currently stated in the User's Manual, we will add this to the manuscript as well.

***

Line by line comments

Lines 9 - 11: "...designed to use either age proxies (e.g., radiocarbon or tephra layers) or stratigraphic alignment (e.g., of benthic δ18O) and cannot combine age inferences from both techniques."

This is a bit misleading because other Bayesian models can indeed use oxygen isotope information -- just not in the way that is done in the present paper. A manually identified oxygen isotope tie point to a reference series can be used together with radiocarbon or tephra, so I suggest this first sentence be reworded.

We will revise this sentence to clarify that this refers to automated alignment.

Lines 28 - 30: "However, this method is restricted to the last 50 ka BP, suffers from variable surface reservoir ages ..., and is often low resolution causing the age model to be highly dependent on assumptions regarding sediment accumulation rate variability."

This is a somewhat outdated viewpoint. An increasing number of labs have installed AMSs, advances in automation have reduced preparation time and cost, and the ability to reliably measure radiocarbon on trace amounts of samples is making it increasingly possible to perform high resolution radiocarbon dating. Also, technically, the current marine calibration curve now extends to 55 ka, not 50 (though I agree that radiocarbon dating at those extreme ages is problematic, and resolution does not match oxygen isotope data). One study that comes to mind is:

Ishiwa, T. et al. Temporal variation in radiocarbon pathways caused by sea-level and tidal changes in the Bonaparte Gulf, northwestern Australia. Quaternary Science Reviews 266, 107079 (2021).

We will revise the sentence to reflect the 55 ka BP end age of Marine20. In addition, we will revise the text to simply state that radiocarbon is generally lower resolution than benthic d18O.

Lines 36 - 37: "Software packages exist to produce probabilistic age models using radiocarbon ages (Blaauw & Christen 2011; Lougheed & Obrachta, 2019), but none of these probabilistically combine age inferences from both dating techniques."

This description should be improved to clarify that the authors are referring to software packages that automatically find the optimal alignment to a reference series. The current descriptions reads as if none of the radiocarbon-centric models can use oxygen isotope tie points.

The sentence will be changed to clarify that it refers to continuous probabilistic alignment rather than tie point identification.

Line 46: Beyond 55 ka

This will be changed from 50 to 55 ka BP.

Lines 48 - 49: "Sedimentation rates are realistically constrained with an empirically derived prior model rather than subjective parameter settings."

I wonder how appropriate this prior is for the possible range of sediment cores that users will inevitably throw at your model? Is it possible to specify your own priors?

It is possible to specify your own prior, which we will specify in the revised text. However, alternate priors should be based on observational data or physically realistic models of the sedimentation process. While some users won't have the interest or specialized knowledge to do this, researchers interested in these processes could publish new prior models that can be used in the software, and we have plans to develop improved priors based on additional data in the future. Another worthwhile feature to implement is the ability to have a prior that depends on the sedimentation environment. Such a prior is conceptually possible with BIGMACS. For now, the manuscript will also be revised to provide more information on the prior we use here from Lin et al., (2014). We will provide core locations and resolutions, and more thoroughly describe our method of construction.

Lines 65 - 66: "Radiocarbon ages must be calibrated from 14C years to calendar years with a calibration curve that accounts for changes in past atmospheric 14C production rates (Reimer et al., 2020; Heaton et al., 2020)"

This is a very simplified statement. Changes in the carbon cycle is also taken into account, and quite a lot of work has gone into better understanding changes in marine reservoir age for Marine20.

This sentence will be revised to reflect the processes described in Heaton et al., 2021 that were considered during the construction of the Marine20 curve.

Lines 67 - 68: "The uncertainty of the calibrated age is a combination of the calibration curve uncertainty, the radiocarbon measurement uncertainty, and the marine reservoir age uncertainty."

To this list should be added 1) local reservoir age offset from the global mean, Delta R, which also has its own uncertainty and is 2) temporally variable (e.g., older water at downstream upwelling sites following AMOC slowdown, etc.)

This will be added.

Line 76: "LGS" should be "LSG". Also perhaps Heaton 2020 should also be cited here because Marine20 includes the BICYCLE LGS-OGCM. (see above comment starting Line 65 -- Both IntCal20 and particularly Marine20 are much more sophisticated (complicated?) than just correcting for changes in production rate.

Yes Heaton et al., 2020 will be cited here and LGS will be corrected to LSG.

Lines 106 and 108: "trial and error"

As with the other models described, Undatable also comes with suggestions regarding parameter selection. Both of these sentences would probably be better without "trial and error". If it took hours to

converge, then that would be "trial and error", but since it takes seconds, It's more like adjusting music volume to one's desired level through instantly received feedback, which is not "trial and error". I'd suggest rewriting as:

"Its quick runtime encourages parameter tuning, based on the authors' recommendations"

and

"These parameters have large effects on the resulting age model requiring the user to decide on the most appropriate values rather than using a prior model of sedimentation rate variability."

And I also suggest that the tunable parameters in the other models be similarly discussed. As it is, this description reads as is undatable is the only one with parameters that can be tuned. This somewhat undersells what the authors are presenting here: a model without tunable parameters.

The tunable parameters in the other models (e.g., the beta and gamma distributions for Bacon and k parameter for Oxcal) are mentioned but perhaps not emphasized as tunable. We will improve the clarity here. We will also make clear that the authors of Undatable offer guidelines when choosing parameter values, and we will remove "trial and error" from the description.

Line 112: "which often correlates with salinity."

yes it might loosely correspond to salinity but it's really surface evaporation - precipitation prior to deepwater formation (since I assume the author's mainly considering benthic oxygen isotopes.

Yes our main consideration is benthic d18O which does often correlate with salinity.

Line 115: "The most conservative technique for aligning records to a target is to assume that large, easily identifiable features in the signals, such as glacial terminations, occurred simultaneously, create tie points between these features, and linearly interpolate between the tie points"

There absolutely is a lag between "upstream" sites in the North Atlantic and "downstream" sites since it can take on the order of 1000 years or more for the signal to propagate with the flow of deepwater.

Yes, we agree that lags exist between benthic d18O records in many cores, and we are not trying to encourage readers to always assume that benthic d18O is synchronous. In fact, in section 4.1, we specifically advise users to align only nearby cores from within the same water mass to reduce the risk of aligning diachronous signals. Here we are trying to give background and context to benthic d18O stratigraphic alignment.

Line 135: I suggest this be better presented with the information starting Line 115.

We state this at the end of the section because it applies to each alignment method. Both tie-point identification and probabilistic alignment suffer from diachronous benthic d18O signals.

Line 165: "termed the likelihoods" remove "the"

We will remove the "the".

Line 171 - 175 The sedimentation model is called a prior distribution which is in turn called a transition model. Perhaps this can be made more clear.

We will reword to state "We refer to the prior for sedimentation rates as the transition model"

Line 180: "confidence" should be "credible"?

Yes. We will change "confidence" to "credible" throughout the manuscript with the explanation upon first mention that "credible intervals" are Bayesian confidence intervals.

Line 190: When are the locations of these 37 cores going to be disclosed?

We will add a thorough description of these cores in the revised draft.

Line 191: "However, where the previous study interpolated sedimentation rates every 1 kyr, we interpolate by 1 cm"

What is the range of sedimentation rates in the 37 cores? Is 1 cm sampling typically equivalent to a 1 ky sampling, or is the interpolation interval vastly different than that used by Lin et al?

We will more thoroughly describe these cores and the calculated sedimentation rates in the revisions.

Line 201 - 202: "Expansion specifies a below average sedimentation rate and refers to a stretching of the local portion of the record."

This is a bit confusing as stated and doesn't become clear until the next sentence where the authors stake that "contraction ... requires squeezing" Maybe rewrite as:

"Expansion refers to a below average sedimentation necessitation stretching the local portion of the record"

Ok we will make this clearer.

Lines 203 - 204: "If the local sedimentation rate is within 8% of the core's average, the state is classified as steady."

How was 8% selected? Please further clarify as is done on lines 209 - 210 regarding the 15 cm interval.

We will more thoroughly describe the transition model and the choices distinguishing expansion, contraction, and steady in the revisions.

Line 220: "improves agreement between the core age models and the radiocarbon observations"

I don't understand this sentence. The age model should be based on the radiocarbon observation in the radiocarbon-dated intervals. Does this indicate that the prior is often vastly different from the data, and

without changing the alpha and beta parameters relative to the previous Bayesian models, the age model obtained by BIGMACS is inconsistent with the radiocarbon dates?

We will rephrase to make our meaning clearer. The likelihood for radiocarbon ages is a student's-t distribution which scales the standard deviation of the radiocarbon measurement. The degree of scaling depends on the values of alpha and beta. Using alpha and beta values of 10 and 11 applies a smaller scaling and a more peaked student's-t distribution that is more similar to the shape of a normal distribution. Thus, age model samples have a higher probability to pass closer to the mean of the radiocarbon measurement compared to using alpha and beta values of 3 and 4.

Lines 234 - 237: "Specifying the model as a uniform distribution will force the age model to pass through the given uncertainty range and should be used when the user is confident about the age information. Specifying a Gaussian distribution will allow the age model to pass farther from the additional age constraint."

This seems backwards to me. If I specify a tephra age as a gaussian distribution with some mean and standard deviation, the highest probability is at the mean, so the model should pass closest to the mean. But if I specify a uniform distribution, the model has an equal probability of passing anywhere. So wouldn't the user want to specify a gaussian when there is good confidence in the age constraint? Perhaps I'm not following what the authors mean to say. Is it that when there is confidence in the *other* age data, with less confidence in the specified tephra/tie point, that the authors are suggesting to use a uniform distribution? I think this statements needs to be clarified.

This is a good point. A comparison between the closeness of an age model with additional age modeled as a Gaussian vs. a Uniform is difficult to generalize for every scenario. We will revise the this statement to reflect this.

Lines 301 - 302: "these cores contain a relatively large number of δ18O outliers (Figure 1)."

Not an appropriate text location to reference fig 1. Please add lat and lon to fig 1.

We will move where we place the call out to Figure 1 and will add lat & long to the figure

Table 1: confirm the longitudes

The longitudes are in degrees East and have been rounded. We will specify °E and provide two decimal places.

Figure 2: a color bar for the panels A and B would be helpful.

Ok we will add a color bar to this figure.

Line 347: "... crosses in Figures 4A and 5A ..."

Figure 3 has yet to be mentioned. Confirm figure numberings. I think this should be Fig 3A and 4A. Generally Figures are numbered in the order they are mentioned in the text.

Thank you for catching this mistake. We will correct the figure numbers.

Line 356: "Figure 6 compares the DNEA and ITWA stacks"

Change to Fig 5.

We will fix this, thank you for catching it.

Line 370: "The Gaussian process regression also creates smoother stacks than previous binning methods"

It would be very useful to the reader to see a comparison of the previous stacking methods. It would also be very helpful to add a figure showing each sediment core's d18O record plotted in a separate panel above the BIGMACS stack. This will let the reader better visualize the the smoothing due to the increased autocorrelation. This would also support the assertion on Line 385 of homogenous signals.

Comparison between the DNEA and ITWA stacks constructed with BIGMACS and the DNA and INA stack constructed with Match from Lisiecki & Stern (2016) is presented in Figure S4. We will add a call out to that figure here. In these plots the apparent smoothing is clear.

While core-specific benthic d18O records are plotted in Figures 3a and 4a, we will add supplemental figures for each core that more clearly plot individual benthic d18O records.

Line 449: "6.1.1 Radiocarbon and multiproxy age models"

Missing from the discussion of applicability is, of course, if the goal is to compare phasing between d18O records, then the multiproxy age model cannot be used and only 14C, tephras, etc. can be used.

This is work we are currently investigating. We have developed a method to calculate lags between benthic d18O records by comparing 14C-only and d18O-only age models. The focus of this section is the advantages of BIGMACS over other age modeling software packages and we will add a description of this application.

Lines 456 - 458: "Because BIGMACS applies a prior model based on observed sedimentation rate variability (Lin et al., 2014), the age uncertainty between 14C observations returned by BIGMACS is physically realistic and less subjective than using tuned parameters in other software packages."

At this point, we still know nothing about the cores from which this prior was obtained. Where are they located? What are their water depths? What are their age ranges? Do they span glacial/interglacial terminations? While this methods does not require parameter selection, it is assuming that the prior is reasonable for the *user's* sediment cores. This is an extremely important point, and I think the authors should spend some time to demonstrate to the reader that the prior is actually appropriate. In short, I'd like to have it explained to me very clearly why the prior assumed here is both appropriate and better than selecting parameters. The statement I mentioned earlier on Line 220 gives the impression that the prior is overly informative and inconsistent with the data.

We will provide a thorough description of the cores included in the prior, the radiocarbon ages, the criteria for inclusion, and the methods to calculate the sedimentation rates. In addition, we will provide a new section in the discussion that addresses the application of this prior on a global scale. In addition, we will make clear that improvement of the prior is currently being investigated. Future versions of the software package will likely have users make a choice between multiple priors based on the sedimentation environment of their core.

Line 470: "widely space"

"spaced"

Thank you we will make this correction.

Lines 478 - 479: "an assessment of a core's absolute age uncertainty should incorporate both the absolute age uncertainty of the target/stack and alignment uncertainty."

I would suggest adding an optional age error column for the stacking target, then fold that error into the alignment uncertainty. You could output both age uncertainty obtained from that of the alignment target, in addition to the alignment uncertainty already returned. The could be added to get a total uncertainty.

This is a potential future revision of the software package, and beyond the scope of the current manuscript. BIGMACS improves upon other age modeling software packages by combining direct dating techniques with probabilistic benthic d18O alignment to construct age models and stacks. During a d18O-only alignment, the uncertainty in the resulting age model only reflects the alignment uncertainty and does not include uncertainty from the target's age model. Combining these uncertainties is an obvious next step. The process of calculating absolute age uncertainty differs for stack construction and benthic d18O alignment and would increase run times. For example, absolute age uncertainty could be calculated for alignments by generating sample alignment targets consistent with the stack/target and re-running the alignment process for each sample target. The computational cost of this calculation is such that all/most currently available automated d18O alignment software (e.g., HMM-Match) does not include this feature. We agree this would be immensely useful and hope to add it to a future version of BIGMACS.

Line 544: "Example multiproxy regional stacks"

The age models are "multiproxy" but the stacks are not.

Thank you we will clarify this.

Line 547: "standard deviations include the effects of spatial variability, age uncertainty ..."

I really think that there should be an easier way for users to include the age uncertainty in the alignment target.

This is definitely something we can address in future iterations of BIGMACS.

Author contributions: It appears that the first two listed authors contributed equally. As such they should be listed as "contributed equally" somewhere around where the corresponding authors are noted. If the other authors only contributed funding for this study, then technically they should not be authors and should be acknowledged.

Yes, the first two authors contributed and equally. We will add this to the corresponding authors list. All other authors directly contributed greatly to this work to and should be listed as co-authors. There contributions are specified accordingly under 'Author Contributions' at the end of the manuscript.

***

I didn't do a full code review but I do have some suggestions as the authors suggest that BIGMACS is resource intensive and slow. There are several things that I see that could be optimized. While I feel that the time and memory savings on the things I am point out will be minimal, it makes me wonder if there are similar inefficiencies in the most critical parts of code.

getInitialTarget.m

Line 66, load calibration curve

Why not just load only the curves that are needed? There are much more efficient ways to read in the data. The fastest is just remove the headerline of each calcurve and use simply load(path). Small things like this, if they occur throughout the code base, can add up to a savings in runtime. Also note that "path" is a command to Get/set search the path. I'd suggest changing the variable name to "Path" or "pth".

The reason why we did not remove the header line is because we hope to make the software more user-friendly: without it, a user might be confused in understanding what each column means. However, your note regarding 'path' is reasonable: we will reflect it to the next version of the software.

Why do all this:

tic

path = 'Defaults/Calibration_Curves/IntCal20.txt';

fileID = fopen(path);

CAL = textscan(fileID,'%s %s %s %s %s');

fclose(fileID);

cal_curve{1} = zeros(length(CAL{1})-1,5);

for k = 1:5

cal_curve{1}(:,k) = str2double(CAL{k}(2:end));

end

toc

Elapsed time is 0.390675 seconds.

when you can do simply this, which is simpler and an order of magnitude faster. Are there similar chunks of inefficient code that are resulting in slow runtime?

tic

Path = 'Defaults/Calibration_Curves/IntCal20.txt';

fileID = fopen(Path);

CAL = textscan(fileID,'%d %d %d %d %d','headerlines',1);

toc

Elapsed time is 0.018334 seconds.

getData.m

If you can figure out what the final size of e.g., "d18O_depth" will be, you can preallocate a matrix for better memory management and faster runtime.

Though loading files is not the bottleneck in cases that we have computed, your comment regarding an efficient coding make sense. We will reflect them to the next version of the software. However, the main limit for runtime efficiency is the use of Matlab. We do not currently have the resources to convert the MATLAB codes into Python or C++, but we hope to do this in the future.

initializeAlignment.m getAlignment.m

This function uses a parfor loop, which requires the parallel computing toolbox that not everyone will have. It will also take time to start up the parallel pool if not already running. Could check for the existence of the toolbox and if it's not installed, use for instead. If there is not a significant improvement in speed, giving the time to start up the pool, it might be better to use just a for loop.

It is true that we have presumed that users may have the parallel computing toolbox and we will clearly state the program requirements in the manuscript.