Response to comments of Anonymous Referee #3

1. General comments

Reviewer comment: In this work the authors present their attempt to harmonize mainly radiocarbon-based chronologies of continental climate records. The harmonisation is with respect of age-model software usage, calibration curve usage, which is a very valuable task. Furthermore, harmonisation is performed with respect to parameters used for the age-depth modelling software. As far as I understand, the authors use the age-modelling software Bacon for age-depth modelling of a huge quantity of records. Before modelling, the cores were manually evaluated in terms of complications, such as radiocarbon reservoir effects, water lines, etc.

While I appreciate their approach, I think there are some things to be improved before suggesting this piece of work for publication.

2. Data (PANGAEA)

Reviewer comment: *(1) Furthermore, I am not able to find age-depth profiles on their provided Pangaea-page. I thought the authors did all their work (handling reservoir effects, water lines, deciding for the best thicknesses to be applied, …) in order to provide a homogeneous age-depth data set. And according to their paper, they spend a lot of efforts to evaluate the datings etc of all records. It would be a pity, if they would not share this. Or is the user supposed to start from scratch again? Even if it 'only' means to run their script – if I understand the code structure correctly, the user has to run all of their thousands of records, even if the user is only interested in one or two records. Especially, as this means to run 'millions of MCMC iterations' (line 120) which cannot be that cheap as even admitted by the authors: "… it needs much supervision and computing power" (line 122). Why not provide all age depth models (including uncertainties) in addition to all meta data and code?*

Response: Thanks for your suggestion. We provided all age-depth models (including uncertainties) on PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.933132; Supplement Table S3) and improved the documentation on the files uploaded in the text.

New text (line 312-324):

'4 Code and data availability

Seven supplementary datasets (Table S1-S7, in comma-separated values format) and one readme text about the LegacyAge 1.0 are accessible in the navigation bar 'Further details' of the PANGAEA page (https://doi.pangaea.de/10.1594/PANGAEA.933132; Li et al., 2021a). We

provided the chronological control points metadata (Table S1), prior information of dates from publication (Table S2), Bacon parameter settings (Table S3), original chronology metadata from the Neotoma and Cao et al. (2013, 2020) (Table S4), LegacyAge 1.0 chronology (Table S5), description of the comparison of original chronology and LegacyAge 1.0 (Table S6), and record references (Table S7) respectively. All datasets are already in long data format that can be joined by the dataset ID.

The R-code for calculation and comparison of chronologies with embedded manual, metadata for code runs, Bacon output graphs of each record, graphs comparison of original chronologies and LegacyAge 1.0, and a short shared-screen video of the R-code to show the usage on two example records are accessible on Zenodo (https://doi.org/10.5281/zenodo.5815192; Li et al., 2021b)'.

Reviewer comment: *(2) Another critical question is about the final age models. As I cannot find them, nor are able to run the R script, I have to ask: Which depths intervals do you choose to save for the homogenised age-depth models? In the paper you mention the effect of choosing different levels or depth intervals on the goodness of the model data and that some are better suited than others. However, I even wonder, why a user should care about having the age-depth relationship on a fixed sampling interval? If I want to work with other paleoclimate data, I am interested in an age-depth model, which provides dates at depth, where the proxies were measured. Unfortunately, this is not mentioned in the paper. Or do you expect the user to apply some (more or less) fancy interpolation algorithm to assign ages for the proxy depths?*

Response: We chose a depth interval of 1 cm to save for the harmonized age-depth models. We reorganized the description of the two Bacon parameters ('thick' and 'd.by') and apologize for the confusion.

'thick' description in the new text (line 167-173):

'(2) Bacon divides a core into many vertical sections of equal thickness (thick; default 5 cm), which significantly affects the flexibility of the age-depth model, and through millions of Markov Chain Monte Carlo iterations estimates the accumulation rate for each section. Blaauw and Christen (2011) indicated that models with few sections tend to show more abrupt changes in accumulation rate, while models with many sections usually appear smoother but are computationally more intense. We run Bacon for six section thicknesses (2.5 cm, 5 cm, 10 cm, 30 sections, 60 sections, and 120 sections), optimal values after numerous tests, with and without core-top age resulting in 12 initial chronologies for each record'.

'd.by' description in the new text (line 182-183):

‘The parameter ‘d.by’ (default 1 cm) defines the depth intervals at which ages are calculated, and we accepted its default value’.

We added a new section in the text to introduce how to assign ages for the proxy depths. For example, we are applying linear interpolation to assign the ages of pollen samples for those records.

New text (line 327-330):

‘5 How to use the LegacyAge 1.0 dataset and code

LegacyAge 1.0 provides the calibrated ages (mean, median, minimum, maximum) and uncertainties at each centimeter for each record with a 95% confidence interval (Supplement Table S5). All users can apply some interpolation algorithms in the chronologies, subsetted from the LegacyAge 1.0 dataset or outputted by our code, to assign ages for proxy depths of records’.

3. Code

Reviewer comment: *(1) Usually, such a data set and code is generated to be used. Unfortunately, I cannot find any description or manual, how to access the age-depth models. Nor is it possible for me to run the R-script. I admit, I am a R-noob, but I think, application should be properly described with at least a short manual for users with some R-experience (or even noobs). This does not have to come with this publication, but it should at least appear on their github space next to the R-file.*

Or at least enable the user to only calculate the age-depth models of the records they are interested in?

Response: We apologize for this. We revised the code and provided the manual and shared-screen video on R-code usage. We also added a description in the text to introduce how to use the code.

New text (line 331-333):

‘As for the R-code, users only need to set the working directory where the Bacon results will be stored and input the record ID of interest to run it successfully. The manual and shared-screen video on R-code usage could provide helpful guidance for users, with or without some R-experience’.

Reviewer comment: *(2) Is the output of your script arranged in a way, that this could be easily accessed?*

Response: We revised the code and provided a manual inside the code. Our code automatically places the different types of files outputted by Bacon in different folders, which will help the users quickly find the files they need.

Manual in the code (line 11-19):

'#-----Resultfolders----------
 # Ages.txt -> Chronology tables by Bacon
 # Bacon.pdf -> Outputplot by Bacon
 # Calibration -> Plots from Calibration
 # ID.Subsets -> Summarized data of the ID
 # Plot.png -> Plot to compare with other chronologies
 # Plot.flipped -> the same plot but flipped
 # Sites -> all data concerning the ID'.

4. Figure

Reviewer comment: *Fig. 7: Please provide information about which of the twelve generated age-depth models for each record you show here! Would it be possible to show one additional age-depth realisation, which fits less good with the measured ages. Only to give the reader an idea about the effects of the choice of depths intervals.*

Response: Thanks for your suggestion. We laid out this figure on a whole page, so we can't show it here. We provided the section thicknesses in the title of each figure and added one additional age-depth realization established by Bacon, which fit less well with the measured ages.

New text (line 296-297):

'Selected typical examples of the comparative results between the accepted LegacyAge 1.0 chronologies, alternative newly generated but rejected chronologies, and the original chronologies are illustrated in Fig. 7'.

Title of Figure 7 in the new text (line 311):

'Figure 7. Comparison of LegacyAge 1.0 chronologies with the original ones. Green line: original chronology. Blue line: LegacyAge 1.0 chronology. Yellow line: alternative newly generated but rejected chronology. Red: date in chronology metadata. Pink: date from prior information. Grey shading: age uncertainties (95% confidence'.

5. Specific comments

Reviewer comment: *(1) L16 and 46: Please elaborate a bit more on what you understand by 'harmonized chronology' already this early in the manuscript. I am pretty, sure, that different people understand different things under this term. I mean later in the paper it becomes clear, what you understand by this term, but I think it is worth to highlight this already in the beginning of your work.*

Response: According to your suggestion, we elaborated this term a bit more in the introduction section.

New text (line 48-51): 'Recently, the need for harmonized and consistent chronologies allowing for the accurate assessment of temporal uncertainty between records has increased as studies are looking for spatiotemporal patterns using multi-record analyses (Jennerjahn et al., 2004; Blaauw et al., 2007; Giesecke et al., 2011; Flantua et al., 2016)'.

Reviewer comment: *(2) L27-28: This sentence needs more explanations. Maybe not here in the abstract, but below in the according text passages. Please find a more detailed comment below.*

Response: Yes, only the final result of the comparison is shown here. We list the criteria below.

New text (line 234-237): 'The criteria for the preferred models are that the model fitted the dates well, had small uncertainties, combined dates with prior information (e.g., geological and hydrological setting, environmental history), and calibrated with the latest calibration curves'.

Reviewer comment: *(3) L69-74: You provide quite some detailed information on metadata, which I appreciate a lot. However, I doubt that putting those data all in one file is the best option. I agree with referee 2 to splitting this file up in several is maybe more appropriate and easier to handle. At least keep this in mind for any potential future improvements.*

Response: We provided seven supplementary datasets in long data format that can be joined by the dataset ID. To avoid misunderstanding, we provide this information now in the text.

New text (line 313-320):

'Seven supplementary datasets (Table S1-S7, in comma-separated values format) and one readme text about the LegacyAge 1.0 are accessible in the navigation bar 'Further details' of the PANGAEA page (https://doi.pangaea.de/10.1594/PANGAEA.933132; Li et al., 2021a). We provided the chronological control points metadata (Table S1), prior information of dates from publication (Table S2), Bacon parameter settings (Table S3), original chronology metadata from the Neotoma and Cao et al. (2013, 2020) (Table S4), LegacyAge 1.0 chronology (Table S5),

description of the comparison of original chronology and LegacyAge 1.0 (Table S6), and record references (Table S7) respectively. All datasets are already in long data format that can be joined by the dataset ID'.

Reviewer comment: *(4) L155: 'acc.mean' is possibly 'acc.rate'?*

Response: The correct abbreviation for mean accumulation rate is 'acc.mean'. We have made the change in the text and apologize for the confusion.

New text (line 161-162): 'The prior for the accumulation rate consists of a gamma distribution with two parameters, mean accumulation rate (acc.mean; default 20 yr cm^{-1}) and accumulation shape (acc.shape; default 1.5)'.

Reviewer comment: *(5) L158: 'We tested six thicknesses (2.5 cm, 5 cm, 10 cm, 30 sections, 60 sections, and 120 sections) …'. I am not very familiar with Bacon. But, why would you want to test those 6 sampling intervals? I mean, the proxies of the cores were measured at specific depths - wouldn't it be more suitable to only interpolate to those depths, where proxy data exist? Actually, this is the data, I would be interested in. But it seems, that this is missing completely. What do you suggest to finally obtain the ages at those depths?*

Response: The selected section thicknesses (2.5 cm, 5 cm, 10 cm, 30 sections, 60 sections, and 120 sections) are the optimal values after numerous tests.

We added a new section in the text to introduce how to assign ages for the proxy depths.

New text (line 326-330):

'5 How to use the LegacyAge 1.0 dataset and code

LegacyAge 1.0 provides the calibrated ages (mean, median, minimum, maximum) and uncertainties at each centimeter for each record with a 95% confidence interval (Supplement Table S5). All users can apply some interpolation algorithms in the chronologies, subsetted from the LegacyAge 1.0 dataset or outputted by our code, to assign ages for proxy depths of records'.

Reviewer comment: *(6) L159: 'artificial surface age', Why would it be necessary to add an artificial date? I don't know if I understand the concept of adding an artificial date correctly. Stating things like this sounds very arbitrary. Or do you mean you added another age-constraint due to the assumption that the core sedimentation was active until core recovery? And that the additional age constraint is the year of core recovery? If yes, please consider to specify accordingly.*

Response: Yes, you are right. To avoid misunderstanding, we rephrased this part of the text.

New text (line 139-143):

'For modern core-tops, if the core was collected from sites where sediment was still accumulating, the sediment surface could be assigned to the year of sampling, adding one significant time control for the chronologies. If the sampling date was unavailable, an alternative surface age from the original chronology in Neotoma was added at the core top. An estimated artificial core-top age (-50 + -30 cal yr BP) was used if none of the above ages were available (Supplement Table S2, S3)'.

Reviewer comment: *(7) L159: 'generating 12 age models for each core'. Just to make sure I understand correctly. Your code provides 12 age-depth models for one core. Are all provided in output files?*

Response: Yes, our code initially outputs 12 age-depth models for each record. We only provided the parameter settings of the 'best' chronology for each record. You will get the best chronology for each record if you run the script directly. Meanwhile, if you want to get multiple age-depth models for each record, you can do so by modifying the column '*Resolution.cm*' or '*Resolution.section*' of Table S3.

Reviewer comment: *(8) L170: I think, C exchange between dissolved C-species in water and atmospheric CO2 is not responsible for 'too old radiocarbon dates'. Instead, this process counter balances to some degree the effect of the arguments listed earlier in this sentence.*

Response: We agree. We rephrased this part of the text to avoid further misunderstandings.

New text (line 187-190):

'Reservoir effects: the uptake of old carbon by aquatic plants, mosses, or shells either originating from, e.g., limestone in the catchment ('hard-water effect') or slow ^{14}C exchange between the atmosphere and ocean interior, can result in too old radiocarbon dates (Philippsen, 2013; Philippsen and Heinemeier, 2013; Giesecke et al., 2014; Heaton et al., 2020)'.

Reviewer comment: *(9) L171-173: For some records you added your evaluation of reservoir effects. I appreciate this a lot, but I think it is worth to add a column in your metadata file and mark those records. This would allow a better transparency about what is your evaluation and which information came from the original studies.*

Response: We listed the reservoir age in column '*Reservoir*' of supplement table S3. Readers can also view 'type 2' (Record with reservoir effect) in column '*Category*' of supplement table S2 to learn how obtained the reservoir effect.

Reviewer comment: *(10) L184: For the use of radiocarbon dates for modelling purposes, you followed 'in most cases the suggestions in the original publications'. Please consider – again for a better transparency - to provide information (maybe in your metadata file), for which records you did not follow the suggestions of the original publications.*

Response: Thanks for your suggestion. We listed all prior information collected from the original publication in supplement Table S2. To avoid misunderstanding, we provided an example in the text.

New text (line 207-209): 'For example, we excluded the date at 164 cm, accepted by the author (Gajewski et al., 2000), from the *Muskoka Lake* record (ID 1783), as it does not agree with the other three dates from the same core and where lithology had changed significantly at that depth'.

Reviewer comment: *(11) L189-191: 'For each record, 12 age models were visually assessed. Preference was given to models that fitted the dates well and with small uncertainties when choosing the 'best' model for each record (Blaauw and Christen, 2011; Blaauw et al., 2018).'. This is a lot of work for thousands of records. You are sure, that you did this all correctly for this large amount of records? I wonder if it would have been more objective to apply a short statistical test on this. I mean, most likely a simple least square test between age model and ages of dated depths would do a better and faster job. Also the 'small uncertainty' argument would be most likely more precise and faster to obtain, when calculating the mean uncertainty instead relying on visual assessment.*

Response: We rephrased this part of the text according to your suggestion.

New text (line 214-227):

'To objectively evaluate the 12 initial age-depth models for each record, we initially tested a least-squares method between the age model and ages of dated depths and calculated the mean uncertainty for each model. However, the least-squares method is susceptible to outliers (Birks et al., 2012), and models with least-squares may risk more abrupt changes in accumulation rate due to over-fitting dates. Instead of a numerical comparison, we finally implemented a visual comparison based on the Bacon output graphs, which show the Markov Chain Monte Carlo iterations, the prior and posterior distributions for the accumulation rate and memory, and how well the model fits the date (Blaauw and Christen, 2011).

Preference was given to models that fitted the dates well, had small mean uncertainties (Supplement Table S5), and good runs of Markov Chain Monte Carlo iterations (i.e., a stationary distribution with little structure among neighboring iterations as indicated by the traceplot of the joint likelihood) when visual choosing the 'best' model for each record (Blaauw and Christen,

2011; Blaauw et al., 2018). If necessary, we adjusted the parameter settings such as the section thickness and mean accumulation rate to better fit with the dates that were consistent with prior information. For the final parameter settings used for each record, please see https://doi.pangaea.de/10.1594/PANGAEA.933132 (Supplement Table S3; Li et al., 2021a)'.

Reviewer comment: *(12) L203: Who did the evaluation about what a reliable date is? You or the original authors? I can imagine, that this is a difficult task, especially for cores from others.*

Response: Sorry for the confusion; we rephrased this part of the text.

New text (line 204-209):

'Dates rejected/added: Neotoma usually reports all ^{14}C dates from cores, even when deemed inaccurate. We assessed prior information on dates and then excluded the ^{14}C dates of samples with contaminated or reworked sediments from age-depth model from age-depth models, in most cases following the suggestions in the original publications. For example, we excluded the date at 164 cm, accepted by the author (Gajewski et al., 2000), from the *Muskoka Lake* record (ID 1783), as it does not agree with the other three dates from the same core and where lithology had changed significantly at that depth'.

Reviewer comment: *(13) L247-248: 'where original chronologies outperformed LegacyAge 1.0, …' How do you know, which model approach outperforms the other? How can you measure or evaluate this? Do you have knowledge of the 'true sedimentation history' of all those records to be able to judge this? Which one do you choose from your 12 ones/core? I think it is very crucial to provide more details on this issue. Or, in case you wanted to express a different thing with this expression, please consider to rephrase this sentence.*

Response: Yes, you are right. We provided the criteria for comparison and rephrased this part of the text.

New text (line 234-237):

'We plotted our newly generated 'best' calibrated chronologies with 95% confidence intervals together with the original ones taken from the Neotoma and Cao et al. (2013, 2020) datasets (Supplement Table S4) to compare and evaluate the performance of the new models visually. The criteria for the preferred models are that the model fitted the dates well, had small uncertainties, combined dates with prior information (e.g., geological and hydrological setting, environmental history), and calibrated with the latest calibration curves'.

New text (line 286-291):

‘For 906 records out of the 2831 records included in the LegacyAge 1.0, no calibrated chronologies were originally available from the Neotoma and Cao et al. (2013, 2020) datasets for comparison. Of the remaining 1925 records, the new LegacyAge 1.0 chronologies were selected instead of the original ones in 95.4% of cases, based on the aforementioned criteria. However, some records still chose the original chronology, mainly because they are varve chronologies, had incomplete metadata (e.g., missing sample depths), or included some non-^{14}C dates that our model could not accommodate (Supplement Table S6)’.