Response to comments of Anonymous Referee #2

1. General comments

Most analyses using Neotoma or other archived pollen data are dependent, at least to some extent, on the chronologies. The available chronologies have variable quality: some record have an uncalibrated chronology, others have a Bayesian chronology. In many cases the uncertainty on the chronology is not available, or if it is, just the upper and lower credibility interval. To synthesise pollen data from several datasets, it may be necessary to harmonised the age-depth models, a huge amount of work. Once such harmonisation is presented in this current manuscript.

(1) As far as I can tell, the chronologies are not archived, but instead the metadata needed to make the chronologies. This is probably a good idea as it encourages the user to check the parameters.

Response: 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). We provided the chronological control points metadata (Table S1), prior information of dates from literature (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. Furthermore, the R-code for calculation and comparison 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.5793936). Thus, readers can obtain chronologies directly using Table S5 we provided or use the script to calculate several or all of the records they are interested in. We also encourage readers to check the parameter settings.

(2) One important result is that "95.4% of records could be improved ". However, it is unclear what objective criteria were used to determine whether the new model represented an improvement. The criteria need to be explicitly stated.

Response: 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, see line 218-225 of new text. Our newly harmonized chronological framework uses the latest calibration curves and Bayesian statistics and fully considers the prior information collected from the original literature.

(3) The metadata and code are available on github (Zenodo.org would be preferable for the final version).

Response: We agree with your suggestion. As we know, Github is the largest and most common source code host. Zenodo provides a persistent DOI to make the work easier to cite, supporting the data from Github repositories. Therefore, we moved to store metadata, code, and results from GitHub to Zenodo (https://doi.org/10.5281/zenodo.5793936).

2. **Data (PANGAEA)**

    (1) The data are arranged in wide format, with a set of columns for each date. This is not the ideal way to arrange the data, as it makes the code much more complicated to deal with this structure, and will need extra extra columns adding in the future to cope with new sites. A better setup would be to have the data in long format, perhaps in multiple files that can be joined by the dataset ID.

    Response: All datasets are already in long data format, i.e., there are fewer variables than observations. For example, there are 23 variables and 19,990 observations in supplement Table S1 at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.933132). While wide data formats take less memory and are easier to analyze, long data formats make adding or subtracting observations easier, as you agreed. If we want to use the ggplot2, an open-source data visualization package for the statistical programming language R, for plotting, the long data format is the best choice.

    (2) At present, datasets are marked as being marine or otherwise. At least in principle, there could be datasets where some dates are on marine fossils, and others on terrestrial macrofossils. Marine should be a property of the date, not the core.

    Response: Yes, you are right. The sites marked as marine or otherwise just distinguish the calibration curve used. Radiocarbon dates of a terrestrial and marine organism of equivalent age have a difference of about 400 radiocarbon years. Terrestrial and marine samples cannot be compared or associated without accounting for the marine radiocarbon reservoir effect. Therefore, we used a different calibration curve for dates on marine fossils than terrestrial. In addition, after rechecking the sites marked as marine in the dataset, all the dating materials are from marine sediments.

    (3) Setting a constant reservoir age for a core without error seems risky to me. Do you check these are only occuring with bulk dates, as there should not normally be a problem with terrestrial macrofossils.

Response: Although the reservoir ages with errors would be better, the values reported by the original author through various methods are mostly constant. We fully respect the original authors' comments because we are no more familiar with the sites than they are. As you can read from the variable 'Material_Dated' in supplement Table S1, there are a variety of dating materials. Aquatic plants, shells, and bulk sediments that may be affected by reservoir effects were often used for radiocarbon dating if ideal dating materials such as terrestrial plant macrofossils have insufficient carbon content for dating. Although the accelerator mass spectrometry (AMS) technique requires much smaller quantities of carbon than the conventional radiocarbon dating, a tiny amount of contamination is sufficient to affect a date. Thus, some dates from terrestrial plant macrofossils mixed with other materials also report radiocarbon reservoir effects.

3. Code

(1) The code is presented in a single script. This is fine if someone wants to make chronologies for all datasets, but often people will want to run a subset of the sites, and may find the script difficult to use. Breaking the script into functions than each do one task, perhaps compiled into a small package with help files and other documentation, would make the product more useful.

Response: Thanks for your suggestion. We reorganized the code and reduced the input files to three tables (Supplement Table S1, S3, and S4) defined in the first 51 rows of code together with an embedded manual. The reader can calculate chronology for all records or some records of interest by modifying lines 35-36. We used URLs for those code calls so that when the code is run, those three input files are imported directly from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.933132). Also, all readers can download these files from PANGAEA or Zenodo (https://doi.org/10.5281/zenodo.5793936) to a new folder and insert the path of the folder to the folder definition at the begin of the code. Additional to the embedded manual, we provided a short shared-screen video in Zenodo to show the usage on two example sites. The embedded manual and the screen video should be helpful as readme/documentation, and now it should be possible to run the code easily. Splitting the code into separate functions is an interesting idea, but getting this to universal usage (now it is very specific and focused on our data) would need more modifications. But to consider this, we separated the "Age Allocation" part into a standalone code that should be the most valuable part of the main script for other purposes.

(2) The code includes fixes for issues the authors found in Neotoma. Please report these issues, and make sure the code to fix them (e.g. line 187) will work safely when the error is fixed.

Response: We excluded the Neotoma downloads from the code because all information needed is located in three tables (Supplement Table S1, S3, and S4), and for them, we can be sure about the data. It would not be possible to reproduce our results with changes in the Neotoma data or the Neotoma package. An automatic modeling with data from Neotoma would be very nice, but without more input information (which we collected in the parameter table) and only standard parameters, the results would not be very satisfying. Actually, the fixes in the code had been implemented just in the case of unit mistakes, and for one site, we contacted the Neotoma team.

4. Figure

(1) Figures 4-6. The x-axis, labelled 0-1, 1-2 etc, is rather cluttered and difficult to read. Please consider other ways to label them. The relative lack of outliers in the pre-LGM sediment is interesting, but probably represents over-fitting the models to sparse data.

Response: The label of the X-axis is changed to not fully displayed, see revised figures 4-5. Skewed distribution of data usually leads to more outliers in the boxplot. As you can read from the following figure (Fig. R1), a time slice includes more outliers if the data are negatively skewed distribution.
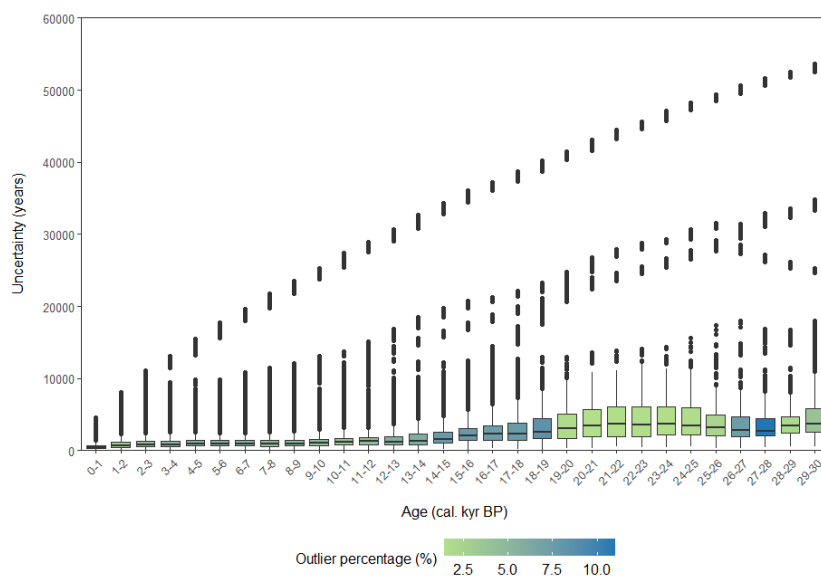

Figure R1. Boxplots of age uncertainties and outlier percentages in distinct time slices (with outliers)

(2) Figure 7 has a minuscule font: I need to view it at 200% before I can read it. No need to repeat the legend four time - once is enough and leaves more space for everything else.

Response: As suggested by reviewer #3, we have newly provided the section thicknesses in the title of each figure, 'best' and 'poor' separately, see revised figure 6. Also, one additional age-depth realization with relatively poor performance has been included for comparison. Preference was given to models that fitted the dates well, had small mean uncertainties, 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 choosing the 'best' model for each record.

5. Specific comments

Line 202 "with fewer than 2" Maybe rewrite as "only one" for clarity. Consideration should be given as to the minimal number of dates that can give a good chronology - I would be cautious using a model based on only two dates.

Response: It has been evaluated that some records do not have one reliable date, and some may have only one, so 'with fewer than 2' may be more appropriate. As we know, the more date a chronology has, the more reliable it is. As you read from column *'Dates_Number'* of the supplement Table S5, 9.7% of the records have only two dates. We have kept these to keep as many records as possible.