Response to comments of Anonymous Referee #1

1. General comments

This paper describes a pollen records dataset, including explanations and descriptions of the dating methods involved in creating the dataset. The global coverage of this dataset is impressive and the presentation of the manuscript is quite good. There are some minor issues with accessing the data, and some considerable issues with the associated code attached to this paper.

(1) While the general shape of the manuscript is good, I encourage a stronger focus on the data itself. These papers are most useful as upfront descriptions of data which requires a slightly different structure than a research articles. Specifically, I would recommend reshaping the intro and the abstract especially to put the data at the forefront, i.e. lead off with statements declaring the dataset, and what it is--for example, putting the name and description of the dataset as the first sentence in both.

Response: Thanks for your suggestion. We have changed the abstract section, i.e., 'We present a chronology framework named LegacyAge 1.0 containing harmonized chronologies for 2831 pollen records…', see line 15-17 of new text. But in the introduction section, we still follow the format of other articles in ESSD, i.e., describing the what, why, and how, to make the reading smoother.

(2) The description of dating methods needs to be expanded briefly, including explicitly defining terms such as "reservoir effect" or clarifying what "insufficient carbon" is. Lead dating is lacking description of methodology as is luminescence. Please also include how these dating methods add to measurement uncertainty in the data. Are uncertainties included?

Response: We have expanded the description of the dating methods, see section 2.2. The uptake of old carbon by aquatic plants or 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, which is called reservoir effect (see line 185-188). Conventional radiocarbon dating requires large amounts of carbon-containing material; otherwise, the dating may not be possible due to insufficient carbon. Furthermore, we have expanded the description of how these dating methods add to measurement uncertainty in the data. The uncertainty of date is available in supplement Table S1 at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.933132).

2. Data (PANGAEA)

This dataset looks to be in good shape and is well-documented when i look at the site the DOI takes me too. When I download the .tab delimited file though, it is really tough to parse. Is there a reason this is in .tab format? A comma separated (.csv format) would be more universal, but I defer to the authors here if there is some subfield specific reason .tab format is better. Admittedly though, I found it difficult to work with this format when downloaded directly. The html web formatted table was easy enough to read.

Response: Seven supplementary datasets (Table S1-S7) 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). As stated in the LegacyAge 1.0_readme.txt, all datasets were uploaded in .csv format, while the default format for PANGAEA storage is .tab. PANGAEA may rename the variables of the uploaded dataset to match its database format. However, these new variable names may have special characters that do not match the requirements of R, so we highly recommended downloading the original file we uploaded before running the R code.

3. Code

(1) The R code that accompanies this data paper and package is highly problematic from an open-code, data sharing perspective. It is formatted for personal use and not up to community standards. The main issue is the beginning call of `rm(list=ls())` This command cleans out and removes all entries in a user's memory and R workspace. Jenny Bryan wrote an excellent piece on why this snippet of code does not work for project based workflows (https://www.tidyverse.org/blog/2017/12/workflow-vs-script/)

Response: As you suggested we removed rm(list=ls()) memory clean. While coding, we were unaware that this could be a problem, so thanks for your input and the link to Jenny Bryans' work. Furthermore, we moved to store metadata, code, and results from GitHub to Zenodo (https://doi.org/10.5281/zenodo.5793936). Zenodo provides a persistent DOI to make the work easier to cite, supporting the data from Github repositories, as supported by referee 2.

(2) The major problem with this becomes apparent a couple of lines down when there are 'fixed' calls to data files that do not exist anywhere--nor can I find them. So running the code is impossible. I would recommend using URLs for those code calls so that when the code is run those data are imported directly from their fixed, online locations. The fixed DOIs from where your data are stored could be used.

Response: Thanks for your suggestion. We 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. 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.

(3) This area of this manuscript/data must be addressed. Additionally, the code is commented adequately, and follows a fairly good syntax, formatting structure. I applaud that. The repo in GitHub though has no readme and no documentation there. That really needs to be added. You could include a lot of what is in this paper, in the data metadata write up elsewhere. I would also encourage including a copy of this manuscript as well as copious amounts of links.

A big ask, which I think would take this next level, is to include a vignette or markdown file showing how to work with his data that includes a small, worked example.

In the current state, I cannot run the code, which gives me pause on my recommendation.

Response: We apologize again for this. Additional to the embedded manual, we provided a short shared-screen video in Zenodo (https://doi.org/10.5281/zenodo.5793936) 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.

4. Specific comments

(1) line 44 - the phrase "calibrated and uncalibrated" is confusing.

Response: Calibrations of radiocarbon age determinations are applied to convert the conventional radiocarbon age to calendar years. Thus, uncalibrated ^{14}C age is the conventional radiocarbon age, calibrated ^{14}C age is the calendar age. We deleted this sentence.

(2) line 65-75 - it would be advisable to have these variables in a table with further descriptions.

Response: Seven supplementary datasets (Table S1-S7) and one readme text about the LegacyAge 1.0 are available at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.933132). Supplement Tables S1 and S4 include all metadata of chronological control points and original chronology from Neotoma and Cao et al. (2013, 2020). Readers can read the readme text to understand the variables better.

(3) line 79-80 - repeated use of references to "most common"

Response: Sorry, this comment is confusing to understand; refer to the repeated use of the word "most common" or the reference? We deleted this word in the new manuscript. But for this reference (Roberts, 2013), various dating methods have been summarized. In expanding the description of the dating methods, we have also added some additional references (see line 150).

(4) Section 2.3.1. - for this type of paper, consider leading this section off with what you have as your final sentence, then describing it. "...all age relationships in our data set are constructed using Bacon..." then describe why and what and how.

Response: As you suggested, we led this section with the final sentence: 'We used the Bacon framework as it is one of the most commonly used methods for age-depth modeling...', see line 120-135.

(5) line 139-141 - where did the latest calibration curves come from? this sentence lacks context.

Response: The latest calibration curves, approved by the radiocarbon community, are stored in Bacon. Readers also can visit them at http://calib.org/. The latest calibration curves (IntCal20, SHcal20, and Marine20) were released in 2020, see line 149-150.

(6) Section 2.3.4 consider laying this section out using bullets or with some kind of work design flow infographic.

Response: We laid this section out using bullets following your suggestion, see line 163-181.

*(7) * just a note format your units with super- and subscripts, not / notation*

Response: We changed '/ ' to superscript ($^{-1}$).

(8) lines 167 -Consider again bullets or something instead of a numbered list inside of a paragraph.

Response: We laid this section out using bullets following your suggestion, same as before, see line 185-206.