Point-by-point changes and responses to comments

Main changes

- Users can now select different calibration curves (IntCal20, Marine20, SHCal20) for their samples
- We reduced the number of sediment cores from 63 to 55, but now only include published datasets
- We compared the LANDO output with previously published age-depth models and adjusted four sediment cores to match the published output
- We changed the legend and plot functionality of LANDO to improve readability

Referee #1 – Bryan Lougheed

<u>Main points</u>

The main issue with the manuscript as it currently stands, in my opinion, is not related to the software itself or how it is described, but how the manuscript uses the LANDO software in an exercise in interpreting the performance of the various age-depth modelling software packages. When comparing the different packages, the authors state:

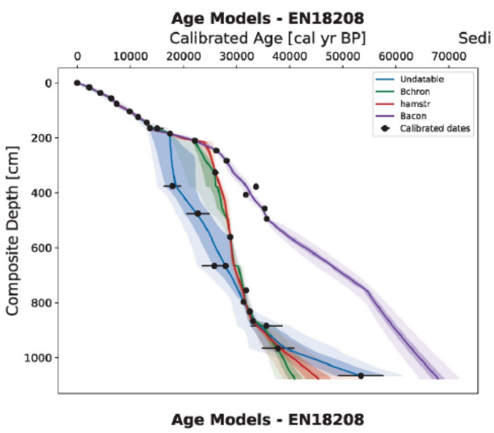
> *"To lower our impact and to avoid introducing biases in the modeling process, we used the default values from each modeling system as our own default values (Blaauw et al., 2021; Blaauw, 2021; Parnell et al., 2008; Dolman, 2021; Lougheed and Obrochta, 2019)."*

The above highlights the general issue with the parts of manuscript that compare age-depth modelling software packages. All of the age-depth model software packages in the manuscript are compared using "default" settings, but all of the packages have settings for a reason, namely that they should be adjusted. So it is possible that the age-depth software packages are not compared on their merits. I note that the LANDO software has the option to adjust the settings for each software package, so I am not describing a limitation of LANDO here.

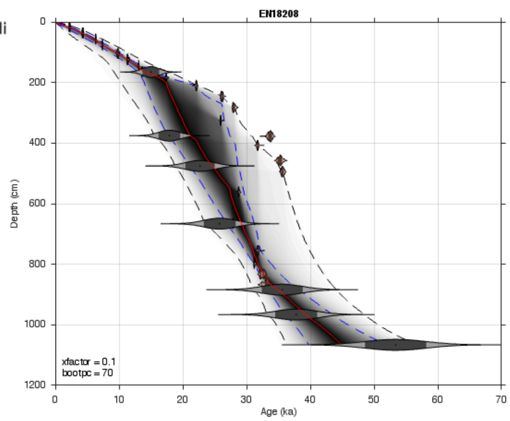
I can give an example about how using "default settings" can affect the interpretation in the case of Undatable. Figure 3 in the Pfalz et al. manuscript suggests that Undatable exclusively follows the younger dates between 200 and 600 cm, and the authors mention something similar in the manuscript in lines 408 to 410 of their manuscript.

While it is true that the GUI version of Undatable displays some settings in the data entry windows when the GUI first boots up, these are by no means "default values", but rather starting/dummy values in the GUI. The Undatable paper (Lougheed and Obrochta, 2019) discusses that bootpc *(bootstrapping percentage) should be increased in the case of large age-depth scatter or age reversals. Indeed, dealing with scatter in this way is stated in Lougheed and Obrochta (2019) as one of the main advantages of Undatable. Seeing as core EN18208 contains such scatter, I have rerun Undatable using a bootpc of 70 (after Gregor Pfalz kindly shared the input data with me), with the following result, with Pfalz et al. Figure 3 shown for comparison:*

Pfalz et al. Figure 3



EN18208 with Undatable, 70% bootstrapping



In the above example, the Undatable uncertainty range expands to take into account the scatter of the dataset, and between 200 and 600 cm the highest probability area shifts more towards the centre of the age-depth scatter. This is the intended philosophy behind the deterministic Undatable, namely that the uncertainty range of the age-depth model should increase so that the scatter of the age-depth points is taken into account, i.e. 95% of the age-depth points should feasibly be located within the 95% uncertainty range of the age-depth model.

Other age-depth modelling packages also have their own settings and approaches.

As discussed in the interactive discussion, we agree with your comment and recognize that this is a broader topic that potentially feeds further research projects related to data science methods. We additionally added to our manuscript the following sentence (Page 25, Lines 559-561): "*For example, to deal with the scatter in the data, users can increase the Undatable parameter "bootpc" to a higher value - as suggested by Lougheed and Obrochta (2019) - to account for a higher uncertainty in the given data.*"

Other points

A small point regarding interpreting a lack of age-depth reversals as "undisturbed sediment"... Following bioturbation theory (e.g. Berger and Heath, 1968) when the sediment is fully uniformly mixed throughout the deposition history, downcore multispecimen / bulk samples will produce *age-depth points that are in chronological order, i.e. lacking age-depth reverals. In other words, a lack of age-depth scatter is not an indicator for undisturbed sediment (despite perhaps 90+% of the literature assuming otherwise).*

Thank you for this comment. We changed the name of the second case study from "Undisturbed sequence" to "Continuously deposited sequence" to ensure that our work also considers for bioturbation theory.

When describing the performance of the age-depth models, the text describes the age-depth models from top down, whereas most of the algorithms operate in the direction of sedimentation/time, i.e. from bottom up.

Thank you – we changed the order in how we describe the performance of the age-depth models in the revised version.

In the age-depth model figures, calibrated dates are indicated by black dots with error bars. Please add some information in the legend or caption about what the black dot is (median or mean calibrated age?), and the error bars (+/- 1sigma, i.e. symmetrical error bars, or the central 68% range, i.e. asymmetrical error bars).

Thank you for the suggestion. We changed the legend of LANDO to use eight different symbols to indicate which category the dating point belongs to. We have updated the figure captions to match their respective symbols and description of error bars.

The "optimised" age-depth model in Fig 4c takes what can be described as a middle route through the age-depth points, but with very small confidence intervals. It could be argued that such small confidence intervals mask the scatter of the age-depth determinations, and therefore the true geochronological uncertainty. This is more of a philosophical point, however, seeing as some age-depth packages try to find an optimised route between age-depth points with minimal age model uncertainty (e.g. Bacon, Bchron, OxCal), whereas others also expand uncertainty to take into account the scatter in age-depth points (e.g. Undatable). An argument can be made for either approach, but in a manuscript that compares all the different types of approaches, it would be useful to point them out.

We agree with your suggestion on pointing out the different approaches. We added a section to the discussion part of the revised manuscript, to ensure that users can decide which approach they prefer (Pages 25-26, Lines 562-569).

Regarding the title, "correlation" would (in my mind, anyway) refer to a statistical relationship of some kind between age and depth. So perhaps replace "correlations" with "relationship" or "models"?

Thank you for the suggestion. We have changed the title to "Improving age-depth relationships by using the LANDO model ensemble".

Referee #2 – Timothy J. Heaton

2 Major Conceptual Comments:

2.1 Suitable Types of Data

Implicitly it seems as though when entering ^{14}C determinations, the data are calibrated against the IntCal20 atmospheric calibration curve - possibly with the application of a reservoir age (although more on that a bit later). Calibration against this IntCal curve is only appropriate for NH atmospheric samples, or for lakes where the reservoir offset is independent of ocean circulation (in such cases the surface water depletion occurs as a potential consequence of both the release of old, but not necessarily dead, organic carbon from soils and peats; and dead inorganic carbon, a hard water effect, entering the lake from its inflows/groundwater). For data from open oceans, one must use the Marine20 calibration curve - it is not appropriate to apply a constant reservoir age to ^{14}C samples from the open oceans since the open ocean environment considerably smooths/filters the (radiocarbon vs calendar age) variations seen in atmospheric signal. Such smoothing does not occur with the application of a constant reservoir age.

However, there are many applications beyond simply lake sediments where one might wish to use age-depth modelling. Age-depth models are frequently used in ocean sediment cores and in archaeological sites. This broadens the potential scope of LANDO. While the introduction discusses only lake sediments, it is not explained as to when the internal calibration process is appropriate and when it is not. Some explanation is needed here as it is likely users will come across LANDO in other contexts beyond simply lake sediments. Further, permittting the users to select the marine calibration curve (with an appropriate ΔR) would increase the applicability of the tool.

I also note that potentially some of the cores, shown in Figure 1 and used in the third case study, look like they might be more general open ocean cores than solely lake sediment. Is this the case? In which case they really should be calibrated using the Marine20 curve - this may also have an effect on the sedimentation rate estimates around the Holocene-Pleistocene boundary since the (open ocean) marine reservoir age is known to change between these two periods - see Figure 4 and 7 in the Marine20 paper.

Again, thank you for your valuable comment. To extend our answer during the interactive discussion, it is now possible to use LANDO for purposes other than lake sediment cores. We state on Page 14, Lines 335- 338: "*To include age determination data within the plots, LANDO internally calibrates the radiocarbon data with the "BchronCalibrate" function of the Bchron package (Haslett and Parnell, 2008; Parnell et al., 2008) with either the*

IntCal20 (Reimer et al., 2020), Marine20 (Heaton et al., 2020), or SHCal20 (Hogg et al., 2020) calibration curve. This allows users to analyze samples from locations other than the terrestrial northern hemisphere." Before calibration, users can also specify the calibration curve for individual samples instead of using one calibration curve for all samples.

2.2 Reservoir Ages

The way that reservoir age is applied in LANDO seems to use a different definition of reservoir age to that commonly in use within the ^{14}C community. For standard IntCal/MarineCal radiocarbon calibration, the reservoir age (at calendar age θ cal yr BP) is defined as the difference between the radiocarbon age of dissolved inorganic carbon (DIC) in the mixed surface layer of the water at that location, and the radiocarbon age of CO_2 in the Northern Hemispheric (NH) atmosphere. In other words, for IntCal and calibration, the reservoir age is measured in ^{14}C yrs and is applied to the ^{14}C determination before calibration.

In this paper however, it seems that the LANDO reservoir age is defined as the difference between the calendar age obtained by calibrating directly against the IntCal curve without any adjustment, and the true calendar age. In the preparation step you use the difference in calendar ages between the unadjusted model and the top of the core.

Applying a constant offset in the ^{14}C domain before calibration of a sample is equivalent to an assumption that a constant proportion of the ^{14}C in that sample arises from inorganic carbon (e.g., the hard water effect). This is not quite true if you simply shift the calendar ages. For old sediment cores (from the pleistocene) or sparsely sampled cores the difference between the approaches may be relatively small - especially for cores that are incredibly long. However I believe this will cause confusion to users.

Thank you for your comment. We assume that there is a slight misunderstanding and hope that our answer in the interactive discussion was sufficient.

2.3 Application on Inconsistent Cores: Example 2

It is my very strong belief that no one should be trying to fit an automated age-depth model to the data as shown in Figure 3. It is clear there is something highly unusual and unexplained regarding the measurements in this core. I am not sure if this is due to certain techniques disagreeing e.g. OSL being different from ^{14}C (since the data are plotted in the same colour and I cannot tell which dates are which). The correct approach would be to go back to the measurements and determine what is going on. In my view, suggesting that data as inconsistent as this can be resolved by forcing them through a range of models (which may or may not happen to select all path through the data) is highly dangerous. This will encourage users to do similarly rather than investigate the root cause of such issues. I believe this case study should not be used for this reason.

In such inconsistent sets of data, I imagine which route the models take through the data will be highly dependent upon their initialisation and, in the case of MCMC, are very unlikely to mix (as can be seen by the narrow uncertainty bands on each individual curve). I do not see this as a case of model averaging (where there are fundamentally different modelling assumptions which are all plausible which lead to different results) but rather luck as to where you initialise each method. In this case perhaps it works ok as the methods happen to choose what look like the most extreme paths however I would think this is to a large extent fortuitous rather than by design.

None of the methods individually fit the data at all well. I would argue that trying to average over a lot of models which are all individually catastrophically bad fits does not add much strength. I much prefer, and would suggest any user takes, the approach of Vyse et al. to go back and look at the raw data and understand what is happening.

Aside: When plotting the data in the cores as solid circles (e.g., Fig 2) I would find it helpful to colour code according to the type of dating used (e.g., OSL, ^{14}C ...). This would make it much easier to identify and understand outliers/inconsistencies. Currently, the text around lines 405 - 410 are not understandable as a reader does not know which are the OSL and which are the ^{14}C dates.

Thank you for this comment. We agree that it is important for any study to look at the data and examine the reason behind the age scattering. We added the following note to our revised manuscript (Page 25, Lines 557-560): *"Although we chose the highest matching score to demonstrate LANDO's ability of filtering out disagreeing models, we do not support the strategy of choosing a single age-depth model with such a low matching score. Rather, users should investigate the cause of the scatter in the age determination data and/or change the default values within LANDO."* We have further improved the plot function in LANDO. Different symbols now indicate the specific material category of the sample.

2.4 Method Description

While I appreciate that the paper is about the development of the coding, I think there needs to be a short description of each age-depth model for the user. This should not repeat the original papers but just give an intuitive explanation so the reader is aware what they are applying, and what the specific assumptions of that technique are. Ideally this section should include an informal discussion of the strengths and weaknesses of each modelling approach. With such a section, a user might be able to decide whether certain models are more appropriate than others for their setting.

Thank you for this suggestion. We extended the part about the age-depth models in the method section by a brief description about the models (Pages 3-4, Lines 108-133).

2.5 Checking convergence

I do not know enough about the specific implementation of the age-depth models in their own packages but is there a way of passing information on model fit to the user. In particular, some of the methods rely upon MCMC and one needs to be assured of convergence; while the frequentist approach (undatable??) may also give some measure of model fit. It may be that the underlying code itself (BACON, BChron, clam, ...) does not provide this but is there a way to obtain information in LANDO that the results of the individual models are appropriate/have converged/or fit?

For example in case study 1, I wondered if it was realistic to have a sedimentation rate that varies from 0.002 to 2.486 cm/yr where the raw ^{14}C dates suggest an inversion? I do not know enough about this location but is it instead possible that some of the models are not fitting well, have been run with inappropriate parameters, or have not converged properly. This is even more *so in case study 2 where I would hope that individually all of the methods would tell you that they are not fitting the data well.*

I recognise this is more about the underlying code to which you link than the LANDO implementation | so there may be nothing you can do to resolve this.

About convergence: Some of the packages provide convergence information in separate methods within each implementation. However, there is no standardized reporting standard for convergence against which we can compare the models. First, we would have to develop such a reporting standard. Then we can build a separate pipeline within LANDO that tracks these convergence values for each sediment core and each modeling system, as our application should continue to work for multiple sediment cores. As this is more a larger feature request, we will address this in a separate update of LANDO. In the initial version of LANDO, there are two extras settings enforced to ensure that models behave appropriately. For Bacon, for instance, we increased the parameter "ssize" as a precaution to ensure good MCMC mixing, as suggested by Blaauw et al. (2021). In clam, we already use the information on the fit to determine the best model for each core, which clam provides as a direct output.

About the example: The values you gave refer to the possible values of the outermost two-sigma ranges when considering all models for the entire core length. Both values do not occur at the same time. We wrote in the revised version (Page 17, Lines 411-412): *"Throughout the core, the cumulative two-sigma uncertainty of the ensemble model ranged from 0.002 cm/yr to 2.486 cm/yr."* At the point of inversion suggested by the calibrates ^{14}C dates, we see high sedimentation rates and only a few models include the dating point at 114.75 cm in their two-sigma uncertainty range.

3 Suggested Additional Information

- *I would suggest you need to provide the link to the software very clearly and explicitly in the Introduction. This is what most users will want and currently you have to get to the end to find out how to actually access the software.*

  Thank you for this suggestion. In addition to the link in the code availability section (according to the GChron manuscript preparation guidelines), we have provided a link to the main repository in the introduction (Page 3, Lines 87-88).

- *I think it worth perhaps adding a clear caveat that it is not appropriate for a user to try all the age-depth models and then simply select the single answer that they like the most in terms of fitting with a particular hypotheses. I know you are not proposing any user do this but a warning might help ensure proper use of LANDO.*

  Good idea. We included the following warning in the discussion section (Page 25, Lines 556-560).

- *You seem to have missed the opportunity to discuss the practical differences between the various age-depth models in your examples. Can you identify features that always seem to be present for some models? For example, in case study 1, the red and green sedimentation estimates seem to be much more extreme than the other models. Is this a consistent feature? Are there reasons for this?*

  We added a description of the practical differences in the method section about the age-depth models (Pages 3-4, Lines 108-133).

- *I do not understand Section 2.5.3 - this needs to be written much more clearly. Does this relate to the way that the proxies and the Holocene boundary are used to filter unreasonable models?*

  We understand that the section might feel a bit out of place. We moved the mentioned section to section number 2.6 and renamed it to "Further analysis - Sedimentation rate development over time". We wrote this section to describe our further analysis that we performed on the multi-core dataset. This section does not apply to the proxies that we use to filter unreasonable models.

- *Important: Table 1 seems to be lacking what I would guess is the most important piece of information - the depth of the measurements. The format of the data on github (and column labelling of the .xls spreadsheet also does not quite correspond to that given in Table 1 - although it is pretty self-explanatory how they transfer).*

  We included the depth information of the measurement in the "measurementID". In Table 1, we define the measurementID as "*Composite key composed of a unique CoreID, a blank space, and* ***the depth below sediment surface (mid-point cm) with max. two decimal digits of corresponding analytical age measuremen****t - example: "CoreA1 100.5", when users obtained sample of CoreA1 between 100 and 101 cm depth*".

- *I think it worth adding a note that all users of LANDO should also reference the underlying methods (and their papers) not simply LANDO.*

  This is a point, which we unfortunately missed. We included the following remark in our revised manuscript (Page 30, Lines 742-744): *"We highly appreciate all the work that went into developing the stand-alone versions of each modeling system. Because LANDO relies on the work of these modeling systems, we encourage users of LANDO to cite the original modeling software alongside the LANDO publication in their work.*"

4 More minor comments

- *I find the use of "correlations" in the title an odd choice. Would it not just be better to have a title "Improving age-depth modelling by using the LANDO model ensemble"?*

Agreed. We changed the title to "Improving age-depth relationships by using the LANDO model ensemble".

- *I agree with one of the comments from CC1 that the parameters chosen for each model are critical. However, I do slightly disagree that this is entirely the responsibility of the user. A reliable method should provide good default values, or ideally an automated method to select good parameters (possibly using e.g. cross-validation??)*

  We agree with both Reviewer No. 1 that parameter selection is crucial and your comment that a reliable method should provide good default values. As we said on Page 27, Lines 632-635 "*But we also wanted to simplify the process for users who do not have in-depth modeling knowledge. By using the default values, we can compare models based on their ability to work with the available data.* ***On the other hand, we are sure that the developers have set their default values based on systematic testing."*** Our aim is to compare age-depth modeling systems given solely the input data. Finding appropriate parameters for each of the individual modeling systems would require techniques, such as grid search, and would be highly dependent on the given input data. This would require direct involvement of the developers of all modeling packages to make suggestions to user on the possible parameters for given input data.

- *Table1 on pg4 - What is the relevance of thickness? It seems unclear to me what thickness of the sample layer means, where is the actual sample within the specific layer? If it is an average over the entire layer then the methodology will presumably become more complex since the ^{14}C determination relates to the average over the respective time period rather than a single calendar year.*

  Thank you for spotting this error. We changed the description of thickness in Table 1 to "***Thickness of the sample slice used for age determination in [cm]***"

- *Should one also have the option of selecting a specific radiocarbon calibration curve? This seems to be particularly relevant for marine ^{14}C samples where one might want to use a Marine curve with a ΔR*

  As we stated before, it is now possible to calibrate dates within LANDO using the different calibration curve (Page 14, Lines 335- 338).

- *Figures lack actual panel labellings such as (a) or (b), e.g., there is no Figure 2a*

  All figures with panels should have a panel label to the right of the sedimentation rate curve. We apologize, if you have received a copy our manuscript without these labels.

- *Line 367'ish - you need to make clear that the mean/median figures relate only to this section (108 - 133 cm in depth) not the entire core.*

Thank you for spotting this inconsistency. We rewrote this sentence to (Page 17, Lines 399-401) "*Mean values ranged from 0.242 cm/yr (hamstr) to 0.764 cm/yr (clam)* ***within this interval****, whereas the median sedimentation rate varied between 0.107 cm/yr (Bacon) and 0.314 cm/yr (clam).*"

- *When plotting the data in the cores as solid circles (e..g Figs 2) I would find it helpful to colour code according to the type of dating used (e.g., OSL, 14C ...). This would make it much easier to identify and understand outliers/inconsistencies. Currently, the text around lines 405 - 410 are not understandable as a reader does not know which are the OSL and which are the ^{14}C dates.*

  Thank you for this suggestion. Since we want to reserve colors for possible additions of age-depth models to LANDO, such as OxCal, we changed the legend of LANDO to use different symbols to indicate which category the dating point belongs to. We hope this will improve the readability of our plots.

- *Line 534'ish | Please identify and explain this unusual core in the section discussing he modelling rather than in the conclusion. I am presuming this is the observation in Figure 6 that has a latitude of around 75° where undatable has a huge difference from Holocene to Pleistocene which is not replicated in the other models.*

  As we explained in the interactive discussion, we mentioned this particular core in the discussion section of the manuscript, but it is now on Page 26, Lines 577-580.

Community comment

(i) Data reference, availability and usage

The origin of the data used in the third case study in most cases is not visible directly from the manuscript but has to be investigated via a "Code and Data availability" spreadsheet, which can be accessed by an attached GitHub link. This spreadsheet in 41 cases provides links to the open and free data repository PANGAEA or original publications, which not always contain the original data and descriptions of age model developments (see below), and in 33 cases the reader is asked to request unpublished data. Once access to the data is accomplished, it is not clear from the manuscript in its present form, which of the existing age data eventually became used in the third case study (see example below).

In our mind the relevant original publications existing have to be cited in the manuscript directly and included in the reference list, the data used in the third case study has to be clarified, and the unpublished data used has to be presented in a table in this paper or at least made freely accessible via an open database.

Thank you for the valuable comment. We agree that it is relevant for the reader to be aware of the underlying data used for our calculations. Fortunately, since the submission of the manuscript, several unpublished datasets have become available in journals. We found only seven unpublished datasets that we excluded from the revised version. In addition, for all of our sediment cores, we referred to the publications with the originally published data and age-depth model. We also qualitatively compared the LANDO model results with the original published age-depth model version and adjusted our LANDO model where needed, for example in the cases listed below.

Following your suggestion, we added a table (Table 4) on data availability within the manuscript containing six columns: "CoreID", "PaleoLake Database ID", "Age-Depth Model Available", "Main Data Source / Repository", "Data Accessible", and "Paper Reference". Furthermore, we included the references given in this table in the main references of the publication. We removed the spreadsheet from the GitHub repository. Instead, we created a table with all dating points including their original reference and submitted the data to Pangaea. This dataset will soon be available to the public.

(ii) Missing geological context

The LANDO-derived sedimentation rates displayed for 39 sediment cores in Figures 5 and S1 suggest continuous sedimentation up to 21 cal ka BP with variable rates. Some of these sedimentation rates are obviously wrong, due to missing consideration of geological evidence. Two examples are given below.

It is true that some of the sedimentation rates do not reflect the actual sedimentation rates compared to age-depth models derived using geological evidence. For this reason, we wrote on Page 26, Lines 570-571 "*Even though LANDO can produce age-depth models for multiple sediment cores ("Multiple cores" – CS3), we must assume limitations in the geoscientific validity for some of the results.*" Since our approach is purely data-driven, i.e., without geological interpretation, we are aware that "*[...] the results from our combined model might over- or underestimate the true sedimentation rate[s]*" (Page 26, Lines 574-575). Our overall purpose was to make LANDO user-friendly enough to allow users to analyze multiple sediment cores without special customizations.

Thanks to your comment, we added the following remark to the manuscript (Page 26, Lines 598-599): "*However, while a specific customization (such as a hiatus) is possible for single core cases, this is not possible in the current version of LANDO for multi-core investigation*."

First, the sedimentation rates derived for core Co1309 from Ladoga Lake are based on age data, which according to the "Code and Data availability" spreadsheet originate from Andreev et al.

(2019) and Savelieva et al. (2019). However, Andreev et al. (2019) only present OSL ages between 118 and 80 ka BP, substantially exceeding the age range of interest here. Savelieva et al. (2019) present the radiocarbon and OSL ages available from the postglacial part of the record, but mention that the age-depth model used originates from Gromig et al. (2019, in Boreas, 48: 330-348), a paper not cited in the manuscript. Gromig et al. (2019) excluded some of the radiocarbon and OSL ages and, on the other hand, added additional age control from varve chronology and correlation with a radiocarbon-dated record close by. Hence, from the references provided it is unclear, which data finally became used for the LANDO calculations presented. Moreover, both Andreev et al. (2019) and Savelieva et al. (2019) mention that the record contains an obvious hiatus, which spans ca. 14-80 ka BP and is described in detail by Gromig et al. (2019). This hiatus is ignored by the LANDO calculations presented, leading to false data at least for the period 21 - 14 ka BP.

We apologize for not including the absolute years of the varve count and the radiocarbon date from Lake Pastorskoye (Subetto et al., 2002), used as an anchor point, published by Gromig et al. (2019) in our original version of the manuscript. We now included data reported by Gromig et al. (2019). However, to allow for a comparison between LANDO and the published age-depth model, instead of modeling the entire core length of 22.7 m, we stopped at the last varve point at 13.23 m. This avoided the problem of extrapolation. As shown in the interactive discussion, LANDO can reproduce the overall age-depth model by Gromig et al. (2019) without removing dating points.

Second, the sedimentation rates presented for core PG1205 from Basalt Lake in East Greenland are based on radiocarbon ages originally published by Wagner et al. (2000 in Palaeo3, 160: 45-68), although reference is made to the PhD thesis of Wagner (2000). The LANDO calculations suggest continuous and relatively constant sedimentation since at least 21 cal. ka BP. However, both Wagner et al. (2000) and Wagner (2000) state that the lake record consists of a till at its base, which in all likelyhood was deposited during the Milne Land stade 11.30 - 11.15 cal. ka BP, overlaid by ca. 6.4 m of glaciolacustrine sediments deposited with high sedimentation rates during deglaciation and ca. 2.6 m of hemipelagic sediments deposited with much lower rates during the past ca. 10 ka BP. Hence, the calculations conducted by Pfalz et al. obviously neglect the regional glacial history presented and discussed by Wagner et al. (2000) and Wagner (2000) as well as many papers published before and afterwards, giving the wrong impression that this part of East Greenland became deglaciated already prior to 21 cal. ka BP.

Agreed. To ensure that we do not include grey literature in the references, we now used Wagner et al., (2000) as reference for the core PG1205. Similar to the Lake Ladoga sediment core Co1309, we changed the length of the sediment core to the last dating point to avoid strong extrapolation in the new version.

These two examples illustrate that neglecting geological evidence for hiatuses or large changes in the rates of deposition can create much larger errors in age-depth models and resulting sedimentation rates than the employment of an age-depth modelling system that may not be ideal for the record investigated. From the two examples it becomes evident to us that the literature existing for all sediment records used in the third case study, not only Co1309 and PG1205, needs to be (re)studied and discussed to assure that the geological evidence provided is considered in the sedimentation rates calculated.

Thank you for bringing this important matter to our attention. We agree and have re-examined all the sediment records closely by comparing the originally published age-depth models and the LANDO outputs. We added an extra paragraph regarding the adjustments on the individual cores (Page 26, Lines 587-601).