Dear Editor,

I appreciate the thoughtful and constructive reviews to my previously submitted manuscript, "Quantifying errors due to discontinuous sampling of paleoclimate archives" (2017PA003246). My replies to these comments are below. Reviewer remarks are shown in grey and additions or modifications to the manuscript are in italics.

In response to the reviews and feedback from others in the community, I have overhauled and broadened the manuscript. The paper is now framed in terms of a new motivating question: how large are "temporal representativeness" errors that arise when one time interval is represented by another? I have connected mathematical results to intuition wherever possible, and have separated the machinery (Section 2) from an application (Section 3) for clarity. The result about time series sampling that was central to the first draft now emerges as a special case of a broader analysis.

Thank you for your consideration of this manuscript.

Dan Amrhein

Reviewer #1

In this manuscript Amrhien discusses the impact of sampling density on aliasing in the context of (sedimentary) paleoclimate archives, which can be a source of uncertainty when attempting to compare data and model output or undertake data assimilation. As mentioned by the author this issue has been raised and discussed in the literature before, although certainly some formalism as provided here as well as a review of best practices would be welcome. From that perspective the subject is interesting.

This said, I found this paper completely incomprehensible. I'm not math-adverse but it was extremely difficult for me to ascertain exactly what the point of this manuscript is.

I have rewritten the paper in order to clarify and expand its objectives. The main objective is to explore how representing one time average by another introduces errors, with potential implications for all paleoclimate archives (not just sediments). The motivation is 1) that previously we lacked a firm basis for how large these errors could be, and 2) the errors can be large in some cases.

To me, Figure 1 tells me that sampling at a low res is find unless you are near the Nyquist frequency, which is something that every paleoclimatologist knows. The power spectra estimate looks fine otherwise. So what do we have to worry about?

As now stated in the discussion (lines 456-458),

When constructing paleoclimate time series, it is important to bear in mind not just the Nyquist frequency but the role of sampling time scales as an anti-aliasing filter...

I have removed the power spectral estimate in order to focus on the mean problem. Figures 5, 6, 7, and 8 demonstrate that error amplitudes can be a large fraction of signal ampiltudes in some cases.

Just glancing at them, the reader has no idea what Figure 3 and Figure 4 mean. This is made worse from the fact that the figure captions are floating in the text and not next to the figures.

I apologize for the floating captions. They have been fixed.

A paper submitted to PP needs to be comprehensible by the PP audience. For that to happen the author has to 1) make it very clear what the problem is and why we need a better solution. Any paleoclimatologist should be able to follow this argument.

In addition to highlighting implications more clearly, I have made a new schematic (Figure 1) that I hope illustrates the potential sources of error that I am trying to characterize. I have also tried to compartmentalize the math and make the implications clear.

2) walk the reader through the appropriate mathematical expressions intuitively. Also, LIMIT how much you show in the main text to avoid equation overload. Put stuff that is ancillary in the Appendix,

Wherever possible, I have included intuitive interpretations of what the equations mean (e.g., lines 206-221). An Appendix contains many of the derivation steps, and I have separated the model construction (Section 2) from the main results (Section 3).

3) Demonstrate clearly how your recommendations improve/solve the proposed problem. The discussion is the place to do this. Give practical examples. How much does this issue matter? Overall: make the manuscript readable even for a geochemist who never thinks about power spectra.

A motivation and summary of results appears in the Discussion, lines 429-453. Having unknown sources of error is problematic, particularly in data assimilation, where uncertainties play a crucial role in balancing contributions from data and models. This framework provides a useful quantification of temporal representativeness error, which turns out to be important in some cases.

Right now all these elements are lacking and a major overhaul is needed. So my recommendation is "reject" but I would encourage the author to re-work this paper and try again. It would probably help to employ a data-based paleoclimatologist as a co-author as they could provide a reality-check on how well the ideas here are communicated, and how important they are. As I said Figure 1 makes me think this is not so important! So better visualization is needed all around.

Feedback from data-based paleoclimatologists helped reframe the paper and hopefully make it more readable and relevant. All figures have been remade to improve communication.

-Line 17: 'growth axis' might make sense for trees and corals but it is a weird term if one is dealing with a sediment core. Use something more general.

I now refer instead to "accumulation axis" for sediment cores.

-Line 70: [complicate] comparisons (?)

This sentence has been removed.

-Line 72: Since this is a single-author paper it should be "I" rather than "we" here and elsewhere.

After considering this suggestion I decided it was best to retain the use of the editorial we, which I (we) intend to include the reader.

-Line 74: Why use TraCE here? Why not use actual Cariaco Basin data?

While my initial goal was to illustrate effects in both data and model output, in the overhauled paper I think this redundancy is not instructive, and have eliminiated the TraCE discussion.

-Lines 78-79: Define what ts and tb are, I'm not sure I understand. From what I can see in the figure you are 1) averaging the record by 100 or 1000 years and then subsampling? If that is the case then there are two effects being tested here together: averaging and subsampling.

A glossary (Table 1) now defines all time scales.

-Line 114: I understand what deltaS is here but what is deltaB? Would this be the bioturbation/mixing depth? Describe what this is in physical terms.

To reduce confusion, I have eliminated the conversion from depth into time, and discuss only time scales; τ_a is now the time scale of archive smoothing.

-Lines 259-261: how so? I'm not following. Example?

A statement on the relevance to data assimilation appears in the Discussion, lines 431-433.

Figures:

-The figure captions are appearing in the text but the figures are at the end which is very distracting and difficult to read. Next time either put the figures in the text also or put the captions next to the figures at the end.

I have fixed this error.

-Figure 2 should be Figure 1 as it is referenced first in the text. Define what all the terms are here in a practical way as it is not clear.

Fixed.

Reviewer #2

In this study, Amrhein investigates the uncertainty associated with the discontinuous sampling of paleoclimate archives and how it might affect model-data comparisons. Although the effort is laudable, I find the manuscript in his current form is lacking background information and a more thorough explanation of the problem this uncertainty poses compared to the largest sources of uncertainty such as those arising from chronologies or calibration to the environmental parameters.

I now include a comparison of errors due to chronological errors (both deterministic, i.e. a known offset, and stochastic, i.e. as arise from radiocarbon dating). Becauses errors from other sources can vary widely in amplitude, and are in some cases poorly known, I have described errors as noise-to-signal ratios.

However, it is important to instill better practices for metadata reporting within the paleoclimate community and this is the perfect example of where additional (easy to report) metadata could improve our understanding of the errors affecting paleoclimate records and facilitate further studies and re-use of the original data. Therefore, the work is important enough for eventual publication in Paleoceanography.

I would recommend addressing the following points before resubmission:

1. Define discontinuous sampling. From Figure2, I'm assuming the author refers to taking discrete samples along an archive with gaps in-between. This type of discontinuous sampling can be remedied by going back and running the analysis on the other samples. However, discontinuous sampling can also occur from nonlinear age-depth relationship. I'm not sure where this problem is take into account in the calculation. Is it taken into consideration with tauS?

I have broadened the paper to focus on the problem of temporal representativity error, of which the discontinuous sampling problem is a part; moreover, I have refrained from using the "discontinuous" terminology, which several people found confusing.

Because all procedures compute errors on a measurement-by-measurement basis, changes in age-depth relationships can be accommodated. The assumption of a locally constant τ_s is made explicit in lines 374-377.

2. Furthermore, how is this problem different from trying to compare a time series sampled at a specific dT (the model resolution) with another sampled with a larger dT?

I am indeed treating these problems as equivalent, as illustrated in the new Figure 3.

Would it be easier to resample the high resolution series to match the dT of the low-resolution time series?

Consistent sampling of records is an effective way to reduce temporal representativity errors. However, if the low-resolution time series in this example were sampled with a short averaging window relative to its dT, then the aliasing errors inherent in this procedure cannot be removed (though they can be estimated).

I do agree with the author that this is only possible is the sampling methods are given in the paper and highlights the importance of proper reporting of metadata in paleoclimate research.

3. The spectral methods employed in the manuscript are for evenly-spaced time series. Do the same bias persist when using methods robust to uneven spacing in X such as the WWZ transform or the Lomb-Periodogram which wouldn't rely on an interpolation and therefore smoothing?

Because errors are estimated on a pointwise basis, there is no restriction to evenly spaced time series. We do make the assumption that the record is approximately evenly spaced near each observation so that we can define a local Nyquist frequency; when this assumption is not valid, the approach will not be accurate (lines 374-377). The potential for using the Lomb-Scargle periodogram or similar measures is referenced in lines 419-421.

4. The motivation of the study is to facilitate model-data comparison. Does applying the correction suggested in the manuscript facilitate this comparison? If so, how? Although I have no doubt that this source of uncertainty becomes relevant as other sources of uncertainties such as chronologies and calibration disappear, I question the usefulness of the correction before these other errors are somehow taken into account.

I now emphasize the importance for model data comparisons by discussing data assimilation in the conclusions (lines 429-437) and acknowledge that TR error may not always be the leading source of uncertainty in the introduction (lines 90-92):

While we do not claim that temporal representativeness error is the most important source of uncertainty in paleoclimate records, it does appear to be large enough to affect results in some cases. Moreover, this work is a step towards reducing the number of "unknown unknowns" in paleoclimate reconstruction.

and the discussion (lines 451-453).

I would argue that a source of error does not have to be the greatest source in order to be worth quantifying. The present subject was chosen (rather than something more proxy-specific) because it is a generic problem that can affect many analyses.

5. Does the author have any idea about the magnitude of the variance introduced by discontinuous sampling vs the other sources of uncertainty?

Because of the many different kinds of uncertainty that affect paleoclimate proxy records, and the fact that those uncertainties are themselves often poorly quantified, I have not included a direct comparison. An exception is errors arising from chronological uncertainty, which are now included in the study (e.g., Figure 7). To give a measure of the relative importance of errors I have compared error variances to the variance of targeted climate signals using the noise-to-signal ratio f. The rationale is that any source of error that accounts for tens of percent or more of the variance in a climate signal is important to quantify.

6. The role of bioturbation: since this is a major source of compensating smoothing effect that is difficult to quantify, I'm not entirely sure how a correction can be made if some of the parameters are completely unknown and expected to vary in time.

Because the methods presented are intended to be used for individual measurements, it is possible to include effects from changing bioturbation or other smoothing. Understanding effects from unknown bioturbation is a possible extension of this model, as noted in the Discussion (lines 473-475):

If time scales (e.g., τ_a) are unknown, a similar procedure can be adopted as was used for Δ in (19), whereby a second integration is performed to compute the expectation over an estimated probability distribution.

7. One of the largest sources of compensating smoothing effects is the water residence time in the karst (speleothem) and trees (water table). How does the karst residence time affects the error due to discontinuous sampling? Especially in relation with hiatuses? Unlike most paleoclimate archives, trees are continuously sampled yet prone to smoothing effects. Does this affect model-data comparison as well?

While the paper does not treat karst or water table residence times specifically, I have made the treatment of smoothing generic to emphasize the many potential sources of smoothing, and include karst as an example (line 135). Effects from smoothing continuously sampled records can be observed e.g. along the diagonals in Figures 8c and 8d; though the relevant time scales for trees are likely to be very different, similar lessons should apply.

Minor Comments: Either attach the caption to the figures or place the figures in the text along with the captions. It's difficult to follow when captions are in line and the corresponding figures at the end of the manuscript.

Fixed.