

Interactive comment on “How large are temporal representativeness errors in paleoclimatology?” **by Daniel E. Amrhein**

Dan Amrhein

amrhein@uw.edu

Received and published: 18 June 2019

The paper by Amrhein addresses an important topic in palaeoclimatology, namely that of time representativeness of proxy data and the subsequent use of that data to compare between models and other data sets. I found the paper very hard to read and very technical, especially for COTP, though I would consider myself at the more technical end of the audience for this type of paper. I wonder if this work might be served better by publishing the mathematics in a more theoretical journal and subsequently interpretation and case studies in COTP.

A key goal of this paper is to communicate some results from time series analysis, which give us a powerful set of tools for thinking about paleo sampling issues, to

C1

the paleo community. I welcome the feedback given on how to streamline the presentation of the mathematical details so that the paper is well-suited to the readership of COTP.

I found the figures well-presented but hard to interpret, and I found the table of notation (Table 1) particularly unhelpful; it looks like it was put together in 5 minutes.

The next draft will clarify figure captions and interpretations.

The key concept seems to revolve around the idea that there is a target measurement x which is desired to be averaged over temporal period τ_x , but the scientist only has access to an observation y which is averaged over an interval of τ_y . The origin of both the target and the observation is the ‘true’ time series $r(t)$. The paper’s goal seems to be to estimate the error in using y to estimate x . I found this slightly strange, for two reasons. The first is that x and y are assumed to be centred at different time points. It’s not clear to me why the centring needs to be (that) different.

I included the possibility of an offset between x and y both to make the problem more general (as such offsets inevitably arise, and I was curious what their effect was) and to permit computing errors from age model uncertainty, which are common in paleoclimate and which are modeled here as a stochastic offset. The zero-offset instance is a specific case of the general treatment.

The second is that, if I were doing this, I would try to estimate $r(t)$ as well as I possibly could, and then create smooths from this to allow me to compare with other time series. Perhaps this isn’t possible, but I couldn’t see why.

It is challenging to estimate $r(t)$ (one has to use paleo data). In particular, we have effectively no information about $r(t)$ at frequencies higher than data sampling

C2

rate, and yet this information is important for estimating representational errors. In the paper I argue that we don't need $r(t)$, but rather an estimate of its statistics, which simplifies the problem somewhat.

I wonder if some of these problems might be solved by having a maximally simple running example at the start that makes clear the novelties of the approach using the bare minimum of maths/notation. That seems to be what has been attempted in Figure 1, though this just raises further questions. The light grey line appears to be $r(t)$ (this isn't mentioned?) which is unobserved. Labelling the vertical axis as $r(t)$ is slightly confusing since y and x are not values from $r(t)$ but averages from it. We would like to get x from y where y is centred on t minus Δ over period τ_y and x is centred on t over period τ_x . The key quantity theta is defined as the different between x and y , and most of the maths proceeds with the assumption that $r(t)$ is weakly stationary.

Thank you for these comments for improving the figure and for the suggestion of a maximally simple running example at the start of the paper. I can see using such an example as a tool for introducing a (reduced) version of the mathematics, and including more of the technical details in the appendix.

At this point I find myself a bit stuck as to the appropriateness of how this works. I don't think I've ever seen a situation similar to this. Aside from the stationarity assumption (which might hold locally?) in all the cases I've worked we have multiple y observations covering the spanned period. (I suppose it might be more realistic if $\tau_x < \tau_y$ as often we're interested in estimating a climate value at a more precise time point than the observations).

I agree that usually more than one observation is being used. One of the motivations for this work, which I can emphasize in the next draft, is this problem of computing mean values from multiple observations. The minimum variance solution for a mean

C3

value depends not just on the values of the multiple observations, but also on their uncertainties (see e.g. Wunsch (2006), 2.7.5). The problem of determining uncertainties individually (what I do in the paper) is thus crucial for determining how measurements should be weighted to estimate time means and uncertainties of paleoclimate intervals.

I'm further confused by the lower panel of Fig 1 which I initially assumed was the distribution of the error described by τ_x (they nearly match in magnitude) but is actually a completely different measurement uncertainty. I don't think that panel is helpful here. However, the legend points out that the uncertainty in theta is characterised by: sampling procedures, variability of r , archive smoothing, and chronological uncertainty. I totally agree with these fantastically important points. It's a strangely important sentence to appear in a figure caption. Having the lower panel display just one of these is confusing.

I definitely see how that can be confusing. The motivation for the lower panel is to illustrate how age model uncertainty can come into play. I will make this point more explicit, and move this key sentence to a more prominent location.

In Section 2 the paper dives into a lot of technical detail which I followed mathematically but quickly lost all the interpretation. I was hoping to pick things up again in Figure 2 which shows the frequency representation of the boxcar function at the top and the power transfer function at the bottom for assumed values of Δ , τ_x , and τ_y . This seems to show which frequencies are contributing most to the TR error. However I've read the bullet points in Section 2.3 about 5 times now and I can't follow them. They talk about certain summary statistics which aren't shown in the Figure (226 and 1325 years?).

Thank you for the feedback. I will move more of the technical description to the appendix to make less of a roadblock for the reader, and will make more explicit

C4

connections between the summary statistics and the figure.

Figure 3 rescues things a little bit by being a bit clearer but even then it doesn't introduce the 4 panels or the colours so I can only make good guesses as to what some these plots are showing. The top two panels are particularly helpful. Unfortunately without any further interpretation of Figures 2 and 3 I was completely lost beyond Section 3.1, which is a big shame. I really wanted to follow this.

I'm glad you found the top two panels helpful, and will make sure that the other panels are explained too (a key point here is about the role of nonstationarity, which you mentioned above).

Subsequent sections explore how errors from different kinds of sampling procedures can build on one another and how they depend on the spectral properties of $r(t)$. I can smooth this transition - emphasizing connections to the simple model and examples - to help the reader connect the dots.

Some minor points:

** I got very confused about the role of τ_a . I thought the archive smoothing was represented by τ_y as that is what we are observing?*

I will clarify this in the next draft. The archive smoothing is meant to stand in for processes (e.g. bioturbation in sediment cores) that can smooth a record prior to sampling it. Sampling also has a smoothing effect, so that a sampled record has potentially been smoothed twice.

** Section 2 I thought was mis-named. It's not really a statistical model in a generative sense. It's more a collection of useful summary statistics that can indicate*

C5

*problems in temporal representativeness. **

I would argue that the model is in fact generative, in that it yields an estimate of errors due to representativeness.

*The glossary in Table 1 needs at least a sentence of interpretation for the more complicated quantities. Saying e.g. $H(v)$ represents the power transfer function is fine, but what would be more useful is to say that high values of this for frequencies v indicate a large contribution to the variance of theta **

Thank you for this suggestion, which I would be happy to implement.

*P6L11 "archives". Also this sentence is quite unclear and where my confusion about τ_a stems from **

I will clarify this in the next draft.

Eq 18 starts to get very confusing when the double square brackets are used to indicate expectations over different random variables. Perhaps use a subscript?

Good suggestion, thank you.

Cited: Wunsch, C. (2006). *Discrete inverse and state estimation problems: with geophysical fluid applications*. Cambridge University Press.

Interactive comment on Clim. Past Discuss., <https://doi.org/10.5194/cp-2019-10>, 2019.

C6