

Interactive comment on “Simple noise estimates and pseudoproxies for the last 21k years” by Oliver Bothe et al.

Anonymous Referee #2

Received and published: 19 December 2018

Review of "Simple noise estimates and pseudoproxies for the last 21k years"

This study presents a generalizable approach to modeling sedimentary proxy systems and then shows how it works using the TraCE-21ka simulation. I think this is a good study that provides a flexible way to estimate various kinds of noise in proxies and that provides a nice set of pseudoproxies for potential use in a pseudo-reconstruction framework. I also think that this study can be useful for seeing how different uncertainties can affect proxy time series.

I have a number of comments, corrections, and requests for clarification below:

Abstract and elsewhere: The use of e.g. and i.e. is too frequent and would be better to just re-write with words.

C1

There are several paragraphs throughout that are just two sentences, which is a little unusual and not totally necessary, and so would be better suited to combine with surrounding paragraphs.

Introduction: Can you better situate the present study in the context of previous approaches to generating sedimentary proxy system models/pseudoproxies? What is unique about this approach? Is it more comprehensive than previous studies? Does it innovatively use the Evans et al. 2013 framework? Is it the first to be applied to the TraCE simulation or to generate pseudoproxies over this time frame? Etc.

p.2 l.8 The words "The review" just after citing both Smerdon 2012 and Mann and Rutherford 2002 make it unclear which paper you're referring to.

p.2 l.12-17 I'm not sure this discussion of "three" different ways is quite right or at least I think I disagree with the framing of the issues here. For instance, the "proxy system model" framework of Evans et al. 2013 subsumes all of these. And so it's not as though using a proxy system model framework is a different approach from just estimating proxy error, it's that just estimating proxy error is usually considering only one of several issues that must be accounted for in the construction of pseudoproxies (i.e., only estimating the "sensor model" while potentially ignoring the "archive model" and the "observation model", using the terminology of Evans et al. 2013).

p.3 l.13-15 It's not clear to me what this sentence means. "On top of this one could use additional stages for the environment and the final reconstruction, however, we can include the associated uncertainties in any of the three stages proposed by Evans et al." The different stages have different types of noise that are particular to the specific process under consideration.

p.7 l.12-15 It's not clear to me what the bias term actually is here. You mention several different things like that it is dependent on insolation, or that it is scaled to be positive, or that it is randomized, or that it is scaled by an ad hoc constant. So what is it then? All of these at once? Only one at a time? Can you state this more clearly and/or perhaps

C2

show in mathematical terms what you mean for the different cases?—just having the term "Bias(t)" isn't exactly clear.

Figure 6: I recommend putting the dates of the periods on this figure so it's more clear which figures correspond to what period (e.g., the deglaciation vs. the Holocene, which have very different correlation maps)

Figure 7: It would help the reader to briefly explain what the values imply. Logs of standard deviation ratios aren't necessarily intuitive. Also indicating the specific date ranges that you're using (as in my comment on Fig. 6) would be helpful.

For both Fig 6 & 7. The color map used here is usually for dry-wet data, but the figures aren't about hydroclimate at all. I would recommend using a different colormap so as to minimize any confusion.

Section 3.5 It would be helpful to write this generalized model down in mathematical terms, not just explain in words, so that readers can be sure what exactly you've done in producing Fig 8 or so that they can think about ways to adjust the generalized model.

p.21 l.11-12 This sentence isn't clear.

Section 3.5.1 Can you motivate the "modifications" you're doing here? It's not obvious to me what needs modification and why. And modifications to which approach, the full version or the generalized one? And what's the motivation for using the generalized approach vs. the full approach?

I think you also need to say more clearly what approach the ensemble of pseudoproxies is based on and why you chose one relative to the other for that dataset. Would it be possible and useful to provide pseudoproxies for both approaches?

p.21 l.14 Why are there 500 pseudoproxies at 144 locations? And are there 500 total or 500 for each of the 144 locations (and thus $n = 500 \times 144$ pseudoproxies)?

Fig 10. The blue lines are hard to see here.

C3

p.23 l.24-25 I'd recommend un-gendering this line using "their"

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2018-137>, 2018.

C4