Clim. Past Discuss., 7, C1422–C1435, 2011
www.clim-past-discuss.net/7/C1422/2011/

Climate of the Past Discussions

Interactive comment on "Exploring errors in paleoclimate proxy reconstructions using Monte Carlo simulations: paleotemperature from mollusk and coral geochemistry" *by* M. Carré et al.

Anonymous Referee #1

Received and published: 19 September 2011

0.0.1 Summary of work.

Carré et al present a series of pseudo-proxy experiments in which they use known "target" SST series to generate surrogate isotopic mollusk shell proxy data, then analyze the uncertainty incurred in reconstructing the target from the pseudo-proxies. Monte Carlo samples are taken from multiple distributions representing various sources of error or uncertainty in the proxy formation process in order to generate ensembles of pseudo-proxy data. This approach provides many replicates or samples of the reconstruction experiments for analysis. The authors are specifically interested in the relative contribution of various sources of error in the proxy data to error in the reconstruction,

and run several Monte Carlo sensitivity tests with various components of error "turned off" so as to isolate the effects of given sources. The algorithm for producing these ensemble experiments, called MoCo, is also presented as a general tool for the paleoceanographic community.

0.0.2 Contribution.

The work provides a well-constructed template for local paleoceanographic pseudo-proxy sensitivity experiments, and so provides a starting point for more quantitative analysis of uncertainty in paleoceanographic studies. Its conceptual simplicity should make it easily accessible to a wide audience. However, the applicability of the algorithm in estimating uncertainties in real-world reconstructions from oceanographic proxies is somewhat limited. In particular, the authors overstate the importance of their algorithm's ability to "[identify and estimate] systematic bias that would not otherwise be detected." MoCo is only able to identify systematic biases in a pseudo-proxy context, where the specific form of climatic nonstationarity over the reconstruction interval is already known. The work also currently does not analyze the effects of proxy uncertainty on any measure of the reconstruction's ability to capture interannual variability. As interannual variability estimates are often the desired product of paleoclimatic reconstructions, the MoCo algorithm seems somewhat incomplete without this analysis.

1 General Comments

1. The authors mention three types of error in the paper (systematic error, potential systematic error, and standard errors), which seem critical to many of their discussions. However, nowhere in the paper are these three types of error clearly and distinctly defined. Clear definitions for these central concepts are critical to the paper, especially since at least one of these labels (standard error) has a different meaning than used here in other contexts (eg. the concept of a “standard error" in statistics).

2. The above comment references the most important example of generally imprecise writing throughout the paper. Explanations of many of the methods need much more careful detailing, and are referenced in the following section listing specific comments.

3. Although the two words are sometimes used interchangeably, the authors may want to reconsider their heavy use of the word “error" throughout the paper and in its title, and replace it with the word “uncertainty." The former word implies a mistake in the analysis, while the latter invokes the effects of an inescapably stochastic and/or nonlinear nature of the proxy formation process on the resulting record of climate. The latter will also tie the work in to a greater body of literature on uncertainty quantification in climate studies.

4. The authors statement of the work's contribution in the abstract and conclusion is falsely inflated. Specifically, the authors need to make it clear in the abstract and conclusion that MoCo is a tool designed specifically for synthetic, pseudo-proxy studies, and that the results mentioned apply only to these contexts.

5. There should also be a section in the body of the paper that clearly and thoroughly addresses the limitations of the algorithm in application to real-world problems.

6. Because an estimate of interannual variability is often the desired product of many paleoclimatic reconstructions, the analysis should look at some measures of the reconstruction's ability to capture interannual variability in addition to the measures already examined (T_m, var(T_m), ΔT, and var(ΔT)). Suggested statistics are the correlation of the estimate with the target, the significance of that correlation, and the coefficient of efficiency and/or reduction of error statistic commonly used in dendrochronology (see Cook and Kairiukstis (1992), eg.). It would also be useful to look at the effects of multiple sources of uncertainty on some measure of variance loss or amplitude attenuation. Analysis of these statistics in addition to the ones already examined will also help place the work in the context of other pseudo-proxy studies in the paleoclimate literature (eg. Smerdon et al (2010), Mann and Rutherford (2002), von Storch et al (2009)).

2 Specific Comments

2.0.3 Abstract

This is well-written and motivated, with the important exception of having swept the pseudo-proxy context under the rug (as noted in the general comments).

2.0.4 1. Introduction

- pp. 2479, lines 16-19: Although I agree broadly with the authors' statement about the assessment of uncertainty in most paleo-oceanographic reconstructions, the authors should be aware of work by Evans et al (1998) on the sensitivity of reconstructions to network choice, and work by Brown et al (2008) and Thompson et al (2011) on the forward-modeling of isotopic signals found in corals.
- pp. 2481, line 2-5: I disagree that the technique presented here is conceptually similar to Haslett et al (2006). The latter study is a climate reconstruction from observed data, and presents modeling that describes the relationship between the proxy signal and the climate, as well as the relationship of the proxy and climate fields to themselves across space and multiple sources of uncertainty inherent in each one. In such Bayesian hierarchical modeling, the data are used to constrain the uncertainties arising at every other level of the model. The work in the present paper is a pseudo-proxy experiment, rather than a reconstruction, and the modeling aims only to represent the effect of various sources of uncertainty in the proxy data.
- In the last paragraph of introduction, the known target and pseudo-proxy context necessary for the operation of the algorithm should be made clear.

2.0.5 2. MoCo Algorithm

2.0.6 2.1

- pp 2482, first paragraph. Looking at effects on metrics of reconstruction of interannual variability would also be useful (for example, correlation and p-value of the reconstruction with target in low and high frequency bands; coefficient of efficiency or reduction of error statistic; reconstruction bias).
- pp. 2482, line 22: The authors describe the proxy formation processes as “stochastic," which may not be strictly accurate (even if inherent nonlinearities in the processes make stochastic models for them appropriate).
- The rest of the section needs to be re-written to clearly define the author's working definitions of “standard error," “systematic error," and “potential systematic error." In Figure 1, definitions for two of the three types of errors are given in equation format; using these equations and elaborating on them in the body of the paper would be useful in providing clear definitions. Note that giving examples of the different error types does not constitute a precise definition.

2.0.7 2.2 and 2.3

- By the end of section 2, the reader should have a clear idea of the general MoCo “workflow." However, I found myself later coming back to this section and comparing it with the details presented in section 4, to try to understand the workflow by example.

  A key point of confusion for me is whether MoCo served only to perturb the climatic target, and requires being coupled to a forward model or proxy formation in order to be used for pseudo-proxy experiments (which seems to be how the case study works). If this is the case, this should be clearly stated in section 2, and the coupling to a separate forward model should be included in the diagram in Figure 1. On the other hand I can also envision the ensemble of perturbed target climates being interpreted as the “pseudo proxy" signal, without the use of a separate forward model, as is currently diagrammed in Figure 1. Perhaps the authors intend for MoCo to be used in either way; whatever the case may be, it should be clearly described here.

If I understand correctly, the intended output of MoCo is a "pseudo proxy", and so should be labeled with a different letter than climate (perhaps use P_i) in Figure 1. These pseudo proxy series must then undergo a reconstruction method before the P_i are translated back into estimates $\hat{C}_i$ of the climate to be compared with the target C_0 (as described later in section 3.4). This reconstruction step should also be described in the section and diagrammed in the figure.

2.0.8 3. Inputs to the algorithm

2.0.9 3.1

- line 13-15: The requirement on the length of the target climate series should be both quantified (longer than the proxy record by what factor? Does that factor depend on any other characteristics of the typical proxy series?) and justified statistically.
- line 17: Such a long time series could also be statistically generated.

2.0.10 3.3.1

- Implicit in the statement that several specimens should be analyzed to average out the effects of spatial heterogeneity is the assumption of some larger-than-

local-scale that the target climate represents. This may not always be the goal of a reconstruction (perhaps one would like to reconstruct climate local to a given proxy), in which case this noise can be set to zero.

2.0.11 3.3.2

- It is not precisely clear what is meant by “proxy analytic error."
- Perhaps explain why these uncertainties add this way to less statistically-oriented readers. For clarity in the equation for σm, either the letters m, a, w, and c should be put in subscript ($\sigma_m^2 = \sigma_a^2 + \sigma_w^2 + \sigma_c^2$) , or else parentheses should be used ($(\sigma m)^2 = (\sigma a)^2 + (\sigma w)^2 + (\sigma c)^2$).
- The choice of temporally uncorrelated errors should be justified for each of these three errors, or perhaps temporally correlated errors should be considered by the authors.

2.0.12 3.3.3

- It should be stated clearly that (Tls, Tli) represent thresholds below and above which precipitation of skeletal material stops. References to the literature to support the existence of these kinds of thresholds should also be provided.
- line 16-18: If these breaks correspond to the input variable “gap" in Table 3, this should be clearly stated.

2.0.13 4: Sensitivity Experiments

- Stochastic noise could also be added to the parameters in model (2) to account for uncertainty in the proxy formation process. Why do the authors choose not to add noise here? (In the language of Bayesian hierarchical modeling, the authors put all the uncertainty at the emphprocess level, and none at the *data level*- this is another reason this study is significantly different than that of Haslett et al (2006)).

- Define V-PDB and V-SMOW before using these acronyms.
- A citation is needed for IMARPE instrumental record mentioned in line 3.
- To clarify the organization of the following experiments, the authors should consider expressing them in terms of statistical factors and treatments. This language could also be incorporated into Table 3.
- The authors may want to consider combining and reorganizing this and the next part of the paper, so that the explanation of each experiment is immediately followed by its results. This would make it much easier to follow.

2.0.14 4.1

- I believe the authors mean to say experiment 1 tests the effect the *number of replicates* or *proxy sample size* had on the standard and systematic errors, rather than the “effect of sampling" or “effect of random sampling". I had to look at the table to clarify what they meant here. This language should be changed for clarity both here and in the results section.

2.0.15 4.5

- How were the two temperature thresholds sampled from the intervals described here? From a uniform distribution on the interval? From a truncated normal distribution? There are many many ways to imagining sampling from this interval, and the authors should be specific about precisely how they did it.

2.0.16 5: Results

2.0.17 5.1

- line 5-7: It is true that systematic reconstruction error increases with the difference between climatic conditions in the reconstructed and calibration intervals. However, it is important to note that this is just one specific manifestation of the general problem that *nonstationarity* poses to climate reconstructions. In real-world reconstructions, there is no calibration test or screening method to detect his kind of error without independent a priori knowledge of the climate outside the calibration interval. MoCo can only detect systematic error resulting from climatic nonstationarity in “pseudo-proxy" experiments.

2.0.18 5.2

- Again, the authors use “random sampling" when I believe they mean to make statements about the effect of sample size on the reconstruction results.
- line 2- 4: The statement comparing errors from short and long records is difficult for the reader to see immediately by comparing Figures 1 and 2. The figures should be made more immediately comparable by plotting N = 200/Ny on the horizontal axes in Figure 2.

2.0.19 5.3

- pp. 2491, line 20: Surely the interpretation should be that the effect of spatial variability on standard error *decreases* with the *number of records* (Central Limit Theorem!), rather than the effect increasing with record length as stated. The authors see an increase with record length only because they keep $N \cdot N_y$ fixed, so N and N_y are inversely proportional. If the authors really want to make statements about the effects of record length, they need to keep the number of records fixed as record length increases so as not to confound influences.

- pp. 2492, line 1- 13: This is very unclear. First of all, if the variability as measured by σ_m can already be translated to temperature variability (as the authors state that σ_m of 0.5‰translates to 2 degrees C), then why bother doing the Monte Carlo simulations? Is the 2 °C before or after aggregating other sources of uncertainty? Clearly I am misunderstanding something about this experiment, and so it needs to be described more explicitly.
- The authors may want to consider making the maximum value of σ_m a fixed factor times the standard deviation of the proxy signal, as in other pseudoproxy experiments (eg. Smerdon et al (2010), Mann et al (2005)) so that the noise is expressed on a scale of signal-to-noise ratio.

2.0.20 5.4

- Discussion of results of experiment 4 are refreshingly clearly interpreted.
- The non-monotonicity of the standard error in Var(Tm) and Var(ΔT) for P.Chicama is quite surprising! This should be explained/interpreted, otherwise this reader is left with suspicion of a bug in the code.
- The asymmetric response to the changing the upper versus lower limit likely has to do with how anomalous the points T_{max}/T_{min} were in the context of the usual climatology. It could be interesting to choose T_{min} and T_{max} based not on the min and max temperatures in the calibration interval series, but on some number of standard deviations above and below the mean temperature.

2.0.21 5.5

- These are good points, but should be discussed in turn as experiments 3, 4 and 5 are discussed, rather than being a separate section.

2.0.22 6: Discussion

2.0.23 6.1

- line 9-10: This is simply the well-known advantage of Monte Carlo techniques that from an ensemble of realizations, it is easy to look at distributions of any function of the reconstruction (eg. distributions of the amplitude, amplitude variance, etc).

2.0.24 6.2

- line 21-22: Citations should be provided for this statement for readers unfamiliar with paleoceanographic studies. It is unclear what is meant by this precision value.

2.0.25 6.3

- In this section yet again, the authors claim that MoCo is an “improvement" over standard reconstruction techniques because it can detect systematic biases, when in fact the algorithm only does this in a pseudo-proxy context. In lines 8-9, the recommendation to seek high temporal variability in target time series seems impractical at best, and fundamentally flawed at worst. In fact the most realistic results will be yielded from a target series that is as *realistic* as possible, rather than as variable as possible.

2.0.26 Figures

- Figure 6: Top panels are unnecessary and confusing; delete them.

3 Typographical Errors and style

- Even though subscripting isn't possible within the body of a code, parameters of the MoCo code, and MoCo input variable names, should be referred to with subscripts in the paper, eg. σ_m and T_{li} instead of σm and Tli. If the inputs and parameters are described clearly enough in the paper, and if the code is commented well, then the correspondence between variable names in the code and in the paper should be clear.
- Figure 5: blue and red are not distinct enough; both look black when paper is viewed at up to 200% magnification.
- pp. 2483, line 19: the word "of" should be the word "one."

References

Brown J, Thudhope A, Collins H M nad McGregor (2008) Mid-holocene enso: Issues in quantitative model-proxy data comparisons. Paleoceanography p PA3202, DOI 10.1029/2007/PA001512

Cook E, Kairiukstis L (eds) (1992) Methods of Dendrochronology: Applications in the Environmental Sciences. Kluwer Academic Publishers, Dordrect, Netherlands

Evans M, Kaplan A, Cane M (1998) Optimal sites for coral-based reconstruction of global sea surface tempeature. Paleoceanography pp 502–516

Haslett J, Whiley M, Bhattacharya S, Salter-Townshend M, Wilson S, Allen J, Huntley B, Mitchell F (2006) Bayesian paleoclimate reconstruction. J Royal Stat Soc: Series A pp 395–438, DOI 10.1111/j.1467-985X.2006.00429.x

Mann M, Rutherford S (2002) Climate reconstruction using 'pseudoproxies'. Geophys Res Lett 29(1501), DOI 10.1029/2001GL014554

Mann M, Rutherford S, Wahl E, Ammann C (2005) Testing the fidelity of methods used in proxy-based reconstructions of past climate. J Clim DOI 10.1175/JCLI3564.1

Smerdon J, Kaplan A, Chang D, Evans M (2010) A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. J Clim 21, in press

Thompson D, Ault T, Evans M, Cole J, Emile-Geay J (2011) Comparison of observed and simulated tropical climate using a forward model of coral δ^{18}O. Geophys Rev Lett DOI 10.1029/2011GL048224

von Storch H, Zorito E, Gonzalez-Rouco F (2009) Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation. Int J Earth Sci 98:67–82, DOI 10.1007/s00531-008-0349-5

Interactive comment on Clim. Past Discuss., 7, 2477, 2011.