Clim. Past Discuss., 5, C758–C767, 2009
www.clim-past-discuss.net/5/C758/2009/
© Author(s) 2009. This work is distributed under
the Creative Commons Attribute 3.0 License.

Climate of the Past Discussions

Interactive comment on "Technical Note: Correcting for signal attenuation from noise: sharpening the focus on past climate" *by* C. M. Ammann et al.

C. M. Ammann et al.

ammann@ucar.edu

Received and published: 10 October 2009

We are very appreciative of the detailed comments on our technical note provided by a total of six referees. The impact of their constructive suggestions will be noticed in the three areas of common focus across the reviews: (1) Improved clarity of the description of the estimation procedure for the unknown variance of the noise, (2) a broader referencing of literature than the narrow focus we intended initially for his technical note, and (3) the further assessment of robustness of the proposed method through inclusion of additional, and memory carrying noise applied to the samples to represent more realistic reconstruction conditions. Because there was substantial overlap in the comments across the reviews (some explicitly recognize that), we provide here a broad summary of these key points first, and then keep the direct replies to points made by individual reviewers brief.

An upfront note to keep in mind when evaluating some revisions provided in the main text as well as the Supplementary Material: This contribution is a "Technical Note" in which we tried to provide a simple description and explanation of the cause of a fundamental problem in regression-based reconstructions that arises from noise in the predictors. Our intent was then to provide an example of how the method works and illustrate its properties through comparison with one simple regression method as a stand-in for the general issues. Neither the actual application, nor the calibration vs verification conditions existing in the real world were of particular concern because the Technical Note is merely an introduction. Its limited space does not provide the opportunity to test the method thoroughly under all possible real-world conditions. However, we do see a point in the suggested limited expansion that reviewers regarded as necessary. We have added simulations with additional noise (both white and red); yet at the same time clearly recognize that these brief tests cannot replace an in-depth, comprehensive evaluation. These need to be done elsewhere, and we invite the community to join in this effort.

We hope that we have answered the questions that relate to the description of the method and the broader context within the existing literature. Additionally, we have expanded on noise-examples that were requested.

Primary points of concern by (essentially) all of reviewers

Clarity of estimation procedure for residual variance: The reviewers correctly recognize that estimating the variance of the noise U is the critical step in our method. As requested, we have expanded the explanation of how a cross-validation step is used to find the minimum bias in the reconstruction of a withheld segment and qualified the statement by acknowledging that other approaches might exist to achieve the same

CPD

5, C758–C767, 2009

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

result. The choice would depend on the data and noise issue at hand. A brief explanation how parameter k is estimated using cross-validation is given in a new first section in supplementary Material. Related to this issue, and only brought up by a couple reviewers is the issue of verification / validation. In real-world application such validation steps are often added to provide measures of confidence in the reconstruction. Here, the focus was on demonstrating the efficiency of the method while the true result is known. Other than the visually clear results, we don't discuss this separate problem in the main text. The issues about verification are independent of the reconstruction method. However, we have included a paragraph with a short discussion in the Supplementary Material and believe that we cover the sentiment indicated by the reviewers. We also reference various discussions in the literature about this very issue that provide further pointers for the reader.

Literature on both broader use of measurement error issues as well as paleoclimate context: We highly appreciate the reviewers' constructive suggestions for references. We are glad to see that there is some awareness of the issues already in the paleo community. This strengthens our confidence that our contribution can have a positive impact and might find interest. As requested, we have broadened our referencing across the board rather than keeping also the reference list as short as possible. There is clear benefit in this, the reviewers are absolutely correct, though we hope the journal can accommodate the number. It is worth mentioning here that we did not reference one particular branch of the literature because we are on record (in publications) for disagreeing with several implementations and conclusions. But given the reviews we decided to provide the full exchanges for the readers. Again, the initial intent was to keep the details as brief as possible for a Technical Note, but recognize that the broader impact benefitted by expanding. We appreciate the insistence of the reviews.

Idealized noise structure in proxy example: The reviewers criticized our overly "idealistic", and in fact "perfect proxy", example. While they are correct that the proxies used for Figure 2 were direct "perfect" samples from a climate model, one needs to recog-

CPD
5, C758–C767, 2009

Interactive Comment

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

nize that twelve points taken from a grid with ∼3600 points does not really present a very “idealistic” exercise. Would we have taken >100 samples as is sometimes used in the literature (the examples the reviewers were referring to), clearly then the reconstruction problem would be different and we would certainly agree with the reviewers. However, using just 12 points for a Northern Hemisphere mean are never considered in “perfect proxy” exercises based on GCM output in the literature. As an argument that the example is not that far away from an actual real world case, we mention in the text that the relationships of the 12 samples to the true (known) NH mean temperature of the model were found to actually be quite comparable to the reconstruction by Hegerl et al. (2006/2007). We still believe that the illustration of the difference between a straightforward Ordinary Least Squares approach and our ACOLS solution is quite striking and drives home the point about attenuation of the signal by noisy predictors.

However, we happily heed to the requests of showing brief examples where “white” as well as “red”-noise were added to the predictors in our reconstruction example. The benefit of ACOLS remains, as expected, yet the improvement for the manuscript is also a more nuanced discussion of how the inflation of variance is distributed in time. Certainly with red-noise present in predictors, the noise and thus the variance in the reconstruction ensemble range gets spread out, just as predicated by several reviewers. We believe that this expanded discussion is indeed beneficial for the paper.

Brief replies to individual requests or criticisms in the reviews:

Ref #1 (Moberg) First point refers to the noise structure used in Figure 2. Moberg suggests adding brief discussion of simulations using different noise characteristics. See response above. Our AR(2) based simulations shown in Figure S3 is demonstrating that ACOLS retains its superior behavior compared to OLS.

Second point (two questions for p. 1648): Explain better, including providing reference to estimate of variance of noise and the parameter k through cross-validation. See also section in Supp. Material. N represents the numbers of years used in calibration.

CPD
5, C758–C767, 2009

Interactive Comment

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

Done. See above.

Third point: Explain error ranges (gray shaded area in Figure 2, and now also Figures S2 and S3): Their computation follows a Monte Carlo simulation (Li et al. 2007, Tellus) and represent 95% confidence intervals based on the full annual resolution that are subsequently (Figure 2, Figures S2 and S3) smoothed with a 10-year Gaussian filter.

Ref #2 (Zorita) Reviewer correctly points out that comparison of ACOLS is done only against OLS. He then provides very valuable background information about nomenclature of different regression techniques and how they are used in different disciplines. This information is important in full-fledged comparisons between methods and would significantly clarify what otherwise remains confusing in collaborative work between statistics and climate disciplines. However, given the brevity of our Technical Note, we decided not to include this rich information. However, we will certainly retain this knowledge and bring it forward in a more in-depth intercomparison of methods, such as the PAGES-CLIVAR Paleoclimate Reconstruction Challenge. We have clarified throughout our paper the naming of what is predictor and what is the predictand.

Zorita also suggests briefly testing the method based on noise-added records so that proxies are more realistically degraded. We have done that (see above). The special calibration issue of 20th century trends is also not explicitly tested here, it already exists in the literature (von Storch et al, 2004; Wahl et al., 2006, Ammann and Wahl, 2007).

Finally, he suggests that calibration statistics with and without ACOLS should be shown. As mentioned above, we have added a new section in the Supplementary Material where we briefly discuss this general issue and provide the important perspective of measuring the quality to reconstruct the mean amplitude. However, verification can be done in many ways and the choice of measure is driven by the question at hand (see Wahl and Ammann, 2007). This problem is independent of the reconstruction method and is necessary in reconstruction intercomparisons, we don't delve into this issue here for space reasons. Here we more simply based our performance evaluation on the skill of reproducing the longterm amplitude.

Abstract: it's not that only the uncertainty bounds increase, the actual variance of the reconstructed series is increased because of the steeper slope. We left the original text.

Text: “uncertainty in the estimation of Y”: done

Figure 2 caption: vertical line separates indeed calibration period: Text adjusted.

3 out of 4 provided references included.

Ref #3 (Christiansen) Reference to Christiansen et al. 2009 now included, but also Rutherford et al. (in review) comment, and the same is true for publication by Buerger and Cubasch. See explanation above.

Include reference Zorita et al. 2003: Done

Improvement of explanation of estimating residual variance: Done, incl. section in Supplementary Material (see above)

Ref #4 (Brohan) He brings up the worry that real world proxy would behave differently than the shown example. We include now white and red-noise examples incl. discussion of the changes in performance that are much closer to real-world conditions (see above), again within the limits available in a Technical Note.

Brohan brings up a “worst” case scenario for paleoclimate reconstructions: Non-linear, or non-stationary climate representation by proxy records. Similar to the discussion of possible various noise-models, we cannot deal with such issues in this Technical Note. Non-stationarity, very likely cannot be dealt with by any method.

Minor Points: 1) New discovery... No, we do not claim this to be a new discovery. Everywhere do we refer to, and even explicitly state, previous literature in the field of statistics. See also abstract. We also point out the broad literature in Astronomy that touches on measurement error. Our goal, and thus the chosen example, is to

CPD
5, C758–C767, 2009

CC BY

provide a simple and concise introduction of the method for paleoclimate purposes. We have improved the referencing (see above) and cover this point quite explicitly. 2) Description of estimation of residual variance (see above) 3) Comparison with other methods: See above. However, Brohan mentions that TLS is not particularly difficult to implement, and estimating two variances is essentially the same as our estimate of sigma_U: We believe that this is not quite the case and actually provide solid literature references by experts in that field of measurement error. Estimating a ratio of variances is inherently volatile. Small errors can have large effects. We point this out by showing the end members of TLS solutions in Figure 1. We do mention that some applications quite likely have this potential error quite small because they could use model data to estimate. But what should be done when noise in observations and in proxies are both unknown? The volatility of the ratio is what led us to ACOLS. 4) Increased variance is actually across all time scales and not concentrated on small scales: We believe that increased variance has different effects. While the full reconstruction is indeed inflated by an increased regression slope, the low-frequency signal is now accurate while the high-frequency signal is potentially quite noisy. By minimizing the bias, we fix the low-frequency problem. In standard OLS, the high-frequency would be quite accurate, but the low-frequency (larger amplitudes) would be off (see bias, and also Fig. S4 in Supplementary Material). Therefore, the selection of procedure will depend strongly on the question at hand.

Ref #5 (Anonymous #3) Issue of “perfect proxy”: see above. This issue is much less severe than the reviewer indicates. We have done the additional simulations for Figure 2. For figure 1 we did not repeat this exercise because the result can be read out of our red-noise example with grid-samples. We keep Figure 1 as clean and simple to understand. The red-noise example is more “applicable” under a real climate problem. Further the surprise for the reviewer that OLS did not behave better is due to the fact that only 12 points were used. This is very different to exercises where >100 samples are taken in the literature (see above).

CPD
5, C758–C767, 2009
Interactive Comment

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

This reviewer also points out the problems of estimating the residual variance properly. We discuss above how we chose our procedure.

Referencing the broader literature, including the trace within paleoclimatology: We have expanded on the references (see discussion above). The specific example about Mann and Rutherford publications not being good representations for attempts to minimize variance loss is actually not quite correct, and we don't' see that our referencing these studies is a mischaracterization. The TLS results (together with the split frequency calibration) did target the amplitude preservation. What was key in these publications was that the underlying data did not contain much more variance! In fact, the coldest proxy conditions occurred around 1900, which is essentially covered by the calibration period. The original method of Mann et al. (1998) however has been found to be potentially biased (lost amplitude) if applied to different data such as GCM data (Ammann and Wahl, 2009). The TLS method later used, however, and thus mentioned as appropriate references, did not substantially suffer from this potential. Therefore, we retain these references. But as the reviewer requests, we also include newer literature more extensively. This should satisfy the reviewer's request for a more broad grounding of the argument in the literature. Other examples given by publications using Borehole data or Mobergs selection of specific low-frequency records are not to the point at hand. The question with ACOLS is if one can better reconstruct the low-frequency signal (mean amplitude) GIVEN data. It is undisputed that should the proxy data not contain low-frequency signals, then even a perfect method could not retain it. Literature: expanded referencing, including most of the requested ones.

Specific comments: 1) Estimating of k through cross-validation: Done, see above and Supp. Mat. 2) "What does insensitive to values around this choice mean": we provide a more explicit range in the text. Done. 3) Variance enhancement concentrated at interannual scale: We have now qualified this statement and if red-noise proxies are used, then the variance inflation will be drawn out to lower frequencies as well. However, we do point out that in contrast to other reconstruction methods, ACOLS does perform

CPD
5, C758–C767, 2009

Interactive Comment

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

very well on interannual scale as well (see Figure S1!). So depending what question is asked, one can comfortably average results to decadal means and talk about climate issues, or retain full annual resolution and look at variability. ACOLS provides both. 4) “Watchful eye on variance”: This expression simply indicates that if very noisy proxies, and particularly if red-noise proxies are used, then the year to year – or even decadal – variance can be substantially larger than true climate variability at that time scale. What our method provides is a robust measure of the mean.

We think we have significantly improved how our text fits into the literature context, provided that one recognizes that this is a technical note to introduce this method in this literature. A full intercomparison is not possible within this framework.

Ref #6 (Anchukaitis) 1) Provides nice literature to measurement error treatment in Astronomy. We believe however that relating past experience of other fields is something more for a review and intercomparison rather than the introduction of a method as a short and concise Techanical Note. We completely acknowledge the need to see the benefits compared to other methods in use, but that also requires full real-world conditions with all its various special issues (incl. trend in calibration period). Therefore, our changes are rather brief pointing more extensively at the literature.

2) Discussion of other regression methods, e.g. indirect versus direct OLS regression. Indeed, each regression method has its strength (or more appropriate physical interpretations). Here we simply can't deal with all options. We tried to stay simple to contrast the non-correction approaches versus a correction due to measurement error through ACOLS. We believe that such requested intercomparisons would be very useful; they simply don't fit into the framework of a technical note. Indeed, when interannual conditions are the focus, ACOLS might not be the optimal solution because of the inflated variance. However, very unusual years such as volcanic cooling years are very well captured because they are dependent on an accurate regression slope. So in such a case ACOLS does provide more faithful estimate. Variability from year to year where the (absolute) mean doesn't matter but the relative changes are of in-

CPD

5, C758–C767, 2009

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

terest, indeed, more traditional methods are likely more successful because they don't experience the inflation of variance.

3) Red noise: implemented, see above. Indeed, we show that the variance is increased under red-noise conditions. This fits the discussion of trade-off between mean bias and variance.

4) Expansion of description of cross-validation step: done, see above.

5) Literature: we now include discussion exchange Mann/Rutherford and Smerdon et al.

6) Title: We agree with Anchukaitis, and have slightly changed the title to : "Technical Note: Correcting for signal attenuation from noisy proxy data in climate reconstructions"

Interactive comment on Clim. Past Discuss., 5, 1645, 2009.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper