Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-211-RC2, 2019

Geoscientific
Model Development
Discussions
Open Access

Interactive comment on "Reconstructing climatic modes of variability from proxy records: sensitivity to the methodological approach" *by* Simon Michel et al.

Anonymous Referee #2

Received and published: 3 February 2019

Review of
RECONSTRUCTING CLIMATIC MODES OF VARIABILITY FROM PROXY RECORDS: SENSITIVITY TO THE METHODOLOGICAL APPROACH
by Michel et al

February 3, 2019

Recommendation: *Major Revisions*

1 Scientific Comments

I'll start with what I like about the paper: it applies several methods to the same dataset, and the results are fairly consistent among methods and with another recent reconstruction, in which one of the authors was involved (Ortegal et al, 2015). That's about it.

1.1 This is no "big data"

Few things are more irritating than people pretending to do "big data" when they actually don't. The authors only end up using a few dozen proxies, and only reconstruct a single index. Nothing wrong with that, but it's not "big data" by any stretch of the imagination. In fact, except for the random forest method (which is only useful in the presence of hundreds or thousands of predictors, therefore not very useful here), all of the methods described are classic forms of linear regression. Anyone is free to call that "machine learning" (since most ML methods are regression in one form or another), but the larger problem is that this is a modeling journal, and I see very little in the way of statistical modeling here.

1.2 Suboptimal Methods

Furthermore, the chosen methods are unable to deal with missing data, forcing the authors to limit the calibration to a set of complete records, thereby jettisoning important information. Meanwhile, at least three methods have been proposed to estimate past climates using discontinuous records:

1. The Expectation-Maximization algorithm (*Dempster et al.*, 1977) and its regularized variants (*Schneider*, 2001; *Guillot et al.*, 2015), as used by *Mann et al.* (2008) to reconstruct the global mean surface temperature, for instance.
2. Bayesian Hierarchical Models, that treat missing observations as extra parameters (*Tingley and Huybers*, 2010a,b; *Tingley et al.*, 2012; *Tingley and Huybers*, 2013; *Barboza et al.*, 2014).
3. Data assimilation approaches, for instance the Last Millennium Reanalysis framework (*Hakim et al.*, 2016; *Singh et al.*, 2018).

All of these methods have code that is publicly archived, often in open-source languages like R.

Restricting themselves to antiquated regression methods forces the authors play a dubious game of optimization on the various training and verification sets, to offset the disadvantage of restricting the network to a gap-less training set. This is suboptimal on methodological and computational grounds.

1.3 How uncertain?

An even more serious issue is that the authors do not provide any measure of uncertainty for their reconstructions. They could do so via any defensible method that has been applied in paleoclimate investigations, e.g. parametric or non-parametric bootstrap, jackknife, or maximum-entropy bootstrap (*Vinod and de Lacalle*, 2009).

1.4 Statistical Models are Models too

I feel compelled to point out that this is a journal about models, so it would be desirable to discuss the advantages of the methodological choices on modeling grounds: each of them models the data and uncertainties in various ways, and it would seem natural for such modeling assumptions and choices to be discussed here (more so than say, Climate of the Past, where the current manuscript would be a better fit in present form). One implicit modeling assumption they make is that the NAO is a linear combination of the proxy data, whereas the correct etiological relationship is the other way around (proxies react to climate, not climate to proxies). This inevitably leads to important biases (*Frost and Thompson*, 2000).

Again, some of the methods mentioned above can deal with that, and the authors should consider using them.

1.5 Perfunctory Validation

Another major problem is that the authors carry out a very perfunctory validation using a metric (correlation) that is known to only reward phase coherence (*Wang et al.*, 2014). At the very least, the authors should explore the Reduction of Error and Coefficient of Efficiently (*Nash and Sutcliffe*, 1970) statistics, which have been used for more than 25 years in the dendrochronological literature (*Cook et al.*, 1994). Another useful measure for point forecasts is the Continuous Ranked Probability Score (*Gneiting and Raftery*, 2007). If the authors were making interval forecasts, which they should, the sharpness of their prediction bands should be evaluated by an Interval Score (*Gneiting and Raftery*, 2007).

Finally, an obligatory measure of any statistical forecasting is to inspect the quality of residuals: since regression relies on residuals being Gaussian, independent and identically distributed, any statistics book (e.g. *Wilks*, 2011) says that the residuals should be tested for these features. This should at least be present in an Appendix.

1.6 Double dipping

The authors pre-screen the proxy network for correlation to the NAO index. What isn't clear is whether that is done as part over the model training, or whether this is done over the entire instrumental era (or the parts of it that overlap with each proxy series). If the latter, this is an example of "double-dipping", whereby information from the test set is used as part of training, leading to overoptimistic results. I could not ascertain this from the paper, so a clarification is necessary.

1.7 Outdated datasets

Why use the PAGES2k version 1, and not PAGES 2k version 2 (*PAGES 2k Consortium*, 2017)? Also, the forcing of *Gao et al.* (2008) is known to contain many errors, which have been corrected by the vastly more complete dataset of *Sigl et al.* (2014). This could explain the very weak signals observed in the paper's Superposed Epoch Analysis. I recommend using the best available data.

2 Editorial Comments

The manuscript reads like a literal translation of a chapter from a French PhD thesis. That means it is 1) overloaded with tedium intended to show that the main author knows what (s)he is talking about; (b) chock full of gallicisms.

2.1 Tedious writing

The description of methods is incredibly tedious. Sections 3.1.2, 3.2.1, 3.3.2 explain the obvious step of linear model prediction as a matrix multiplication. None of this is useful in any way as long as the code is shared. Also, an entire appendix is devoted to a user's guide, which should really be a readme file on GitHub. Please do not waste the readers' disk space and printer ink with this.

One of the most tedious parts is that the *PAGES 2k Consortium* (2013) paper is consistently referred to as "the Pages 2K database 2014 version". Since it was published in 2013, why insist on calling it 2014? Also, the consortium's name is "PAGES 2k", not Pages 2K.

In section 3.1.3, several approaches are mentioned to choose the truncation parameter (none, it should be said, with the aid of any statistical theory), but they are not used. Either leave them unsaid, or mention them and use them (e.g. by comparing what choice is obtained with those methods vs cross-validation).

2.2 Gallicisms

The manuscript is generally well organized, but the writing suffers from many gallicisms. Since I happen to know a little French, here is an attempt at translating them:

- page 6, line 11: facilitate → simplify
- page 8, line 11: most performant → best-performing
- page 11, line 16: inversed → inverted
- page 12, line 23: to present frequently a → to often result in a
- page 15, line 15: require to be tuned → require tuning

2.3 Unavailability

I understand the need to protect data and code until the paper is published. However, acting like they are public, and linking to a non-functional Zenodo link (https://zenodo.org/record/1403146#.W4UMUGaB2qA) is bad form. Either give a complete link or mention that the data/code will be shared upon publication.

References

Barboza, L., B. Li, M. P. Tingley, and F. G. Viens (2014), Reconstructing past temperatures from natural proxies and estimated climate forcings using short- and long-memory models, *Ann. Appl. Stat.*, *8*(4), 1966–2001, doi:10.1214/14-AOAS785.

Cook, E. R., K. R. Briffa, and P. D. Jones (1994), Spatial regression methods in dendroclimatology: A review and comparison of two techniques, *International Journal of Climatology*, *14*, 379–402, doi:10.1002/joc.3370140404.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion), *J. Roy. Stat. Soc. B.*, *39*, 1–38.

Frost, C., and S. G. Thompson (2000), Correcting for regression dilution bias: comparison of methods for a single predictor variable, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *163*(2), 173–189, doi:10.1111/1467-985X.00164.

Gao, C., A. Robock, and C. Ammann (2008), Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models, *Journal of Geophysical Research (Atmospheres)*, *113*, 23,111–+, doi:10.1029/2008JD010239.

Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, *102*(477), 359–378, doi:10.1198/016214506000001437.

Guillot, D., B. Rajaratnam, and J. Emile-Geay (2015), Statistical paleoclimate reconstructions via Markov random fields, *Ann. Applied. Statist.*, pp. 324–352, doi:10.1214/14-AOAS794.

Hakim, G. J., J. Emile-Geay, E. J. Steig, D. Noone, D. M. Anderson, R. Tardif, N. Steiger, and W. A. Perkins (2016), The last millennium climate reanalysis project: Framework and first results, *Journal of Geophysical Research: Atmospheres*, *121*, 6745 – 6764, doi:10.1002/2016JD024751.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proceedings of the National Academy of Sciences*, *105*(36), 13,252–13,257, doi:10.1073/pnas.0805721105.

Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, *10*, 282–290, doi:10.1016/0022-1694(70)90255-6.

PAGES 2k Consortium (2013), Continental-scale temperature variability during the past two millennia, *Nature Geosci*, *6*(5), 339–346, doi:10.1038/ngeo1797.

PAGES 2k Consortium (2017), A global multiproxy database for temperature reconstructions of the Common Era, *Scientific Data*, *4*, 170,088 EP, doi:10.1038/sdata.2017.88.

Schneider, T. (2001), Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values., *J. Clim.*, *14*(5), 853–871, doi:10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2.

Sigl, M., et al. (2014), Insights from antarctica on volcanic forcing during the common era, *Nature Clim. Change*, *4*(8), 693–697.

Singh, H. K. A., G. J. Hakim, R. Tardif, J. Emile-Geay, and D. C. Noone (2018), Insights into atlantic multidecadal variability using the last millennium reanalysis framework, *Climate of the Past*, *14*(2), 157–174, doi:10.5194/cp-14-157-2018.

Tingley, M. P., and P. Huybers (2010a), A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 1: Development and applications to paleoclimate reconstruction problems, *J. Clim.*, *23*, 2759–2781, doi:10.1175/2009JCLI3016.1.

Tingley, M. P., and P. Huybers (2010b), A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 2: Comparison with the Regularized Expectation-Maximization Algorithm, *J. Clim.*, *23*, 2782–2800, doi:2009JCLI3016.1.

Tingley, M. P., and P. Huybers (2013), Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, *Nature*, *496*(7444), 201–205, doi:10.1038/nature11969.

Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam (2012), Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quaternary Science Reviews*, *35*(0), 1 – 22, doi:10.1016/j.quascirev.2012.01.012.

Vinod, H. D., and J. L. de Lacalle (2009), Maximum entropy bootstrap for time series: The meboot r package, *Journal of Statistical Software*, *29*(5), 1–19.

Wang, J., J. Emile-Geay, D. Guillot, J. E. Smerdon, and B. Rajaratnam (2014), Evaluating climate field reconstruction techniques using improved emulations of real-world conditions, *Climate of the Past*, *10*(1), 1–19, doi:10.5194/cp-10-1-2014.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences: an Introduction*, 676 pp., Academic Press, San Diego.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-211, 2018.