Clim. Past Discuss.,
https://doi.org/10.5194/cp-2018-1-RC2, 2018

Climate of the Past Discussions

Interactive comment on "Autoregressive Statistical Modeling of a Peru Margin Multi-Proxy Holocene Record Shows Correlation Not Causation, Flickering Regimes and Persistence" *by* Seonmin Ahn et al.

Anonymous Referee #2

Received and published: 30 April 2018

The article aims at presenting a new way of investigating the causal relationship between different proxies extracted from the same core through the analysis of conditional probabilities in an autoregressive model with two states. I particularly enjoyed that the authors made their software public on Github.
I found that the article was quite "technical" and the journal "Climate of the Past" is probably not a convenient choice. Despite the fact that the approach is original and potentially useful to paleoceanographers, I think the authors should review the article and go deeper into the statistical analyses before resubmitting the manuscript. I also

Printer-friendly version

Discussion paper

suggest the authors to think about which journal to (re)submit in. I would have opted for a journal like "Nonlinear processes in geophysics".
Please find below the main points which I think should be improved. As I am a theoretical climatologist, I focus here on the methodology.

1. Basic correlation analysis.
   As it is mentioned in the title and the abstract, the study aims at showing that this new method is better than basic correlations when we want to explain causal relationships. I suggest to add a full paragraph where you show the classical correlation analysis and point out its main disadvantages, in order to better motivate the subsequent techniques.

2. In the introduction, you should explicitly state that the four proxies come from the same core, as it is an important implicit hypothesis.

3. In the introduction, the choice for the age model should be better discussed. In particular, you should discuss the changes you expect if you work with state-of-the-art age models based on Bayesian statistics, from a the technical point of view (is it technically feasible to fully consider the probability densities in your framework?) and from a climatological point of view (is it worthwhile?).

4. In section 2.3, you mention that considering more than two states basically gives the same results. I suggest to show it in an appendix. Moreover, you should first study the one-state case and show its limitations.

5. On the model.
   I am quite familiar with such equations but it took me a lot of time to understand your model. You should better explain that it is the same equation for the two states but with different parameters. Writing explicitly the HMM case would also make things clear. As the HMM is not well-known in the climate community, I

suggest you to briefly review the technique in an appendix jointly with the Baum-Welch algorithm, and in particular the concept of "hidden variable".
For a given state, the model is basically a 4-D Gaussian white noise (HMM) or a 4-D Gaussian red noise or AR-1 (AR-HMM). This should be explicitly mentioned. It turns out that the estimation of the coefficients of such processes can be performed from irregularly sampled time series, i.e. it does not require the data to be interpolated, as you do through the Kalman filter. See e.g. Gardiner, 2009, Stochastic methods, 4th edition, Sect. 4.5.6, and Kelly et al., 2014, the astrophysical journal, 788: 33. Transferring these techniques to your case is certainly not direct, but I think it is worthwhile investigating it to get rid of the interpolation procedure, which may cause uncontrolled biases in the analysis, especially if you want other people to use your code on data with any possible sampling schemes.

6. In section 3.3, you conclude that the four proxies are not causally related. I think you should better motivate this by a proper statistical test rather than just looking at the numbers in the tables presented at the end of the article.

7. In Fig. 2, 3 and 4, please draw in function of the time/age rather than the time step.

8. In Fig. 5, you should indicate state 1 for the top figures and state 2 for the bottom ones.

9. In table 1, you should point out and explain why the 4-4 matrices are symmetric.

10. In tables 3 and 4, please provide the explicit formulas for the computation of those matrices.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2018-1, 2018.