Thanks to the authors for revising the manuscript and responding to my comments. The manuscript is very much improved, but I have a further comment to their response, followed by some small remarks on the revised manuscript.

Regarding comment 2 and L572-581 in the revised manuscript.

What I was proposing is to use a mixture distribution. The individual DO events are then regarded as subpopulations. Allowing for different subpopulations accounts for the fact that the DO events are heterogeneous, and thus cannot be expected to all show the exact same value of time lag in between proxies, just as the authors are arguing. Still, they are all DO events and thus it makes sense to ask questions on the population level, which is described by the mixture distribution. So one can e.g. simply test whether the population mean is greater or equal zero.

As I said previously, to me this is the obvious alternative to what Erhardt et al were doing. They are multiplying the individual posteriors, while in a mixture distribution the individual posteriors are added, thus yielding a larger variance. This increased uncertainty would most likely yield a non-significant lag in the proxies, in line with what is presented in the paper.

I understand from the viewpoint of the authors why they want to construct a 16-dimensional distribution. But I am not sure why they invoke “physical quantities”. In my opinion this is just a matter of different interpretations of uncertainty, and I would like to hear why the authors think what they are doing is more physical. The way that the authors argued in their response so far is by assuming a priori that their interpretation is the only correct one. I do agree that the two procedures do not necessarily give the exact same results, but I would just like to hear from the authors why they think their complicated approach is necessary.

I guess the main difference to just using a mixture is that the procedure presented here accounts for the fact that we have just 16 observations, which gives additional uncertainty, as reflected e.g. by the larger standard deviation in their example. This information is lost when one just looks at the mixture distribution. But it can be easily recovered, e.g. when using a bootstrap test with the given sample size.

Related to this, in my earlier comments with bootstrapping I meant simply random sampling. A mixture distribution is constructed by random sampling from the individual components, and that’s why I was thinking that this approach is very similar to the authors’.

Again, I am not saying what the authors are doing is wrong in any way, or identical to what I am proposing. The manuscript would simply benefit if the authors could clarify why not a much simpler, and equally reasonable, approach is taken.
At least statements like “*the combined set of MCMC samples does not correspond to any mathematical object and hence its interpretation remains unclear.*” in the revised manuscript need to be corrected.

L86ff: “*In order to review the statistical evidence for a potential systematic lags, we formalize the notion of a ’systematic lag’: We call a lag systematic if it is enshrined in the random experiment in form of a population mean different from zero. Samples generated from such a*

biased population would systematically (and not by chance) exhibit sample means different from zero. Accordingly, we
formulate the null hypothesis of a pairwise unbiased transition sequence, that is, a population mean equal to zero."

Could this maybe be reformulated, in order to also acknowledge the fact that whether or not a truly biased sample can be expected to systematically exhibit sample means (significantly) different from zero depends on the sample size? To me, this seems to be an important motivation for the procedure proposed here.

Caption Fig. 2: I was not sure at first whether the posteriors shown are from the authors or from Erhardt et al. Maybe this could be made specifically clear in the caption.

L187: indipendent → independent

L797-798: I don't understand what is meant here, maybe just a grammatical error?