Review of

# Patterns of changing surface climate variability from the Last Glacial Maximum to present in transient model simulations

by Ziegler et al.

**Recommendation**: *Accept with major revisions*

Summary The manuscript presents a detailed comparison of transient model simulations since the last Ice Age, using a hierarchy of physics-based 2D or 3D models, from simple energy balance models to Earth system models. The work is rigorous, well-motivated and well-written, and I am generally supportive of publication once a few key issues have been addressed.

## 1 Scientific Comments

### 1.1 Death by moments

I appreciate the authors' effort in setting up a general framework by which to analyze this ensemble of simulations. However, having to compare 4 moments across 15 simulations makes for a very large volume of information, bordering on overwhelming (witness the 41 figures in the supplement). This would be warranted if there was significant dynamical/physical insight to be gained from this assessment, but I found the paper lacking in this regard. For instance, what do changes in kurtosis (of temperature or precipitation) tell us about? Can this be tied to particular dynamics (e.g. convective clouds vs stratiform clouds produce precipitation distributions that can be distinguished by these moments). If not, is this really useful? In many places the insistence on painstakingly documenting the minute details of all four moments in these models makes for very bulky prose, from which this reader derived very few insights (e.g. L543-555). A statement like "The chosen ice sheet reconstruction has a limited impact on temperature kurtosis on all timescales analyzed here" does not help explain any dynamics. It also seems (e.g. L503) that it is broadly insensitive to many factors, so perhaps it is not a very useful indicator of anything? If so, why fill the paper with it?

I recommend using higher-order moments only when they can be connected to identifiable processes, and/or if they can be constrained by paleoclimate observations (proxies). Otherwise, this reads like a gratuitous exercise that multiplies figures without simultaneously enriching the content.

### 1.2 Chapter or Paper?

The methods' description, as well as the background, read more like a thesis chapter than a paper in a specialized journal, where some common understanding exists. Recalling the mathematical definition of the various moments (section 3.1), or of spectral quantities (section 3.2), seems overkill here. Either readers know it, or they can look up a standard stats textbook. Similarly, the opening

paragraph of section 3 is appropriate for a thesis chapter, but superfluous in a paleoclimate journal, where everyone either already understands this mathematical framework, or doesn't care enough about it for this exposition to matter. I recommend stripping down this exposition to the bare minimum, particularly if the code will be made available. The authors only need describe what they did in very general terms, and interested readers can go look at the code if they want to reproduce any of the results.

## 2 Editorial Comments

**L25** Insert full stop before "rather" and start a new sentence there.

**L114** "has been shown to reduce seasonal to interannual standard deviation" missing "the" before "seasonal"

**149** "normality assumptions might break down" -> this runs counter to my experience, whereby on long timescales, any averaging process (which is common in climate) makes things more Gaussian, by virtue of the central limit theorem. Can the authors explain why they expect normality to break down here?

**L173-174** : "They vary regarding simulation setup, applied forcings and model complexity." This is redundant with the previous paragraph.

**L189** "present-day": please define. The present is a singularly ambiguous notion in paleoclimatology [*Wolff*, 2007].

**L219** "2 kyr AP". Presumably "AP" stands for "after present", but since the present was not defined, this does not help. Also, this is the first time I see this acronym, so the authors should add a footnote to explain what they mean.

**Fig 2, caption** "with respect to the past 2 kyr": Is this the reference for everything? If so, please state explicitly in the text, upstream of this caption, so there is no ambiguity.

**Fig 3a** this is a very creative and helpful way to represent model complexity and help compare the various models used here. Most of the dimensions are qualitative, but atmospheric and oceanic resolution are two dimensions where one could be quantitative. Can the authors at least give a sense of the range of resolution spanned by the model ensemble?

**L296** “and EBM” : suggest "and *the* EBM", since there is only one here.

**L312** "from the end of the LGM to PI". Is PI the same as "the present"? See comments above, and please be consistent in what baseline is used for the modern era.

**L317-321** superflous paragraph

**L325** : "we remove the trend from the timeseries using a Gaussian filter" -> presumably this is a high-pass filter? Note that for discrete data a binomial filter is more justified.

**Sections 3.1, 3.2** tighten up so they read more like a paper and less like a dissertation chapter.

**L398** "a frequency range of [$2t_s$,1000]" -> technically, this is a range of periods

**L411** "Overall, MPI-ESM r1âĂŞr7 exhibit the largest temperature difference" Can this be tied to the equilibrium climate sensitivity of the various models?

L422]: "standard deviation provides a measure that increases with the spread of the distribution" -> technically, is IS a measure of the spread of a distribution. I find this wording needlessly mathematical.

**L437** : "On centennial scales, this lack of skewness agrees with the results for the LGMR." It should be pointed out that LGMR uses a Kalman Filter that assumes Gaussian state vectors, so no skewness could be reconstructed, even if it was there.

**Fig 5** : would it make sense for the $y$ axes to have units?

**L465** "standard deviation changes only locally" -> What does this mean? There are definitely some continental or basin scale patterns in some of the panels.

**L485** "differences in parametrization" influence the skewness. Which parameterizations might be responsible here? It may not be easy to guess, but if the authors have candidates in mind, they would be helpful to lay out here.

**L540; L554** : the word "significant" is used on those 2 lines. Do the authors mean "statistically significant" (if so, by which test?) or do they mean something like "substantial"? Please clarify.

**Fig 12** is very dense, and too small to distinguish many of the curves. The 6 PSD panels on the left have a curve/envelope in light gray, but there is no corresponding entry in the legend.

**L588** "we find temperature spectra that increase towards longer timescales" -> this certainly appears to be the case. Can the authors be more quantitative here? What scaling exponents are involved? How does it compare to observationally-derived exponents? Is there evidence of multifractality? If so, are the regime transitions occurring at the right timescales?

**L594** "relates to the simulated El Niño-Southern Oscillation". This seems to imply that those simulations exhibit enhanced ENSO activity with LGM boundary conditions. Is there any published explanation for this?

**L602-603** : again, this can be quantified with scaling exponents.

**L625** : "into the tropic" -> into the tropics

**section 4.6.4** : LGMR provides an ensemble of Kalman Filter samples. Is this taken into account here, or are the authors only considering the ensemble median?

**L672/3** : "Simulations that differ only by ice sheet reconstruction diverge most on long timescales, although differences can be found even for annual variability". This suggests that those boundary conditions affect the entire climate continuum, which is profound and deserves some commentary.

**Fig 13** : it took me a minute to figure out that the two gray lines correspond to PMIP end-members. The authors might want to clarify this in the figure caption.

**L758-759** : "Further, as a reanalysis product, the LGMR uses model simulations as priors and thus might be affected by a lack of variability in models." This intuition is incorrect. In the offline data assimilation flavor of ensemble Kalman FIlters used in products like LGMR or LMR, all temporal variability comes from the proxies; the model priors are used only to link

variables across space, or to link one variable to another (e.g. surface temperature to sea-level pressure). So this comment would only be true if it applies to spatial variability.

**last sentence** reads very generic. Please make more substantive, or dispense with it.

## References

Wolff, E. W. (2007), When is the "present"?, *Quaternary Science Reviews*, *26*(25), 3023–3024, doi: 10.1016/j.quascirev.2007.10.008.