Review of the article "Deciphering the spatio-temporal complexity of climate change of the last deglaciation: a model analysis"
by D. M. Roche, H. Renssen, D. Paillard
Clim. Past Discuss. 6, 2593-2623, 2010

In this manuscript the authors discuss a fundamental problem in paleoclimatology: the temporal sequence of global and regional climate changes during the last deglaciation process and the lead and lag relationship among forcing and regional changes.

The authors made use of an Earth System Model Intermediate Complexity, LOVECLIM, and simulated the last glacial termination from 21,000 BP to 9,000 BP by prescribing orbital forcing, greenhouse gas concentrations, and forcing from the ice-sheets. Then, they applied a formal statistical test procedure to identify, where and when the first significant changes in the mean temperatures and precipitation can be detected.

Key results from this model analysis are that sea-ice covered polar latitudes respond first to orbital and greenhouse gas forcing, and that the timing depends on the season. Furthermore, they identify regions where one cannot detect significant changes throughout the entire deglaciation process, which provide guidance in the selection of paleo-proxy sites for the study of deglaciation.

The paper is exclusively focused on the model results, however, the major conclusions are also relevant for the analysis and interpretation of paleo proxy records.

In general, the manuscript is well-written and presents an important contribution the ongoing debate of the lead and lags between various proxy records, and the causes for the last deglaciation.

Scientific significance: good

Scientific quality: good

Presentation quality: good

Does the paper address relevant scientific questions within the scope of CP?
yes

Does the paper present novel concepts, ideas, tools, or data?
yes

Are substantial conclusions reached?

yes

Are the scientific methods and assumptions valid and clearly outlined?

yes (with some minor improvements suggested)

Are the results sufficient to support the interpretations and conclusions?
The model results and statistical tests are sufficient, however, some extended discussion might be needed regarding the interpretation of the statistical tests (see below).

Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?
yes

Do the authors give proper credit to related work and clearly indicate their own new/original contribution?
yes (few citations suggested below)

Does the title clearly reflect the contents of the paper?
yes

Does the abstract provide a concise and complete summary?
yes

Is the overall presentation well structured and clear?
yes

Is the language fluent and precise?
yes

Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?
(yes, but see suggestion below)

Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
Yes, the discussion of the statistical test results.

Are the number and quality of references appropriate?
yes

Is the amount and quality of supplementary material appropriate? NA

Individual comments:

Abstract:
p.2594,line 5: write: [...] focused on the understanding of the complex sequence [...]
p.2594 l. 11: write: [...] we do not include freshwater forcing [...]
p.2594 l. 14: write: [...] forcing, and the locations wheres those [...]

1 Introduction:

p.2594, l.23-25: add references (e.g. Hays et al., Science, 194, 1976; EPICA, Nature, 429, 2004; Huybers, Nature, 434, 2005)
p.2594-2595, l. 25-/ l.2): please restructure statement (Milankovitch did not have the information from the paleo records at that time). suggested: "Milutin Milankovitch was one of the first to propose that low-frequency variability of the orbital parameters, which modifies the energy received at the top of the atmosphere, could be the cause for glaciations on Earth."
p.2595 l. 6 : write: [...] (LGM) ice sheets covered [...]

p.2595 l. 14: [...] deglaciation have arisen [...]

p.2595 l.17-19 suggested: [...] there is still some debate on how changes relate to each other at different geographical locations on Earth (Stott et al, Science, 2007; Huybers and Denton, Nature Geoscience, 2008;Timmermann et al, Paleoceanography, 2009; Denton et al., Science, 2010). Though such [...]
p.2595 l. 23-24: write; [...] within the physical processes contained [...]

2.1 Model description:

p. 2596, l. 15-17: mention that CLIO uses rotated grid.

2.2 Deglacial forcing:

p.2597, l.4 : "Our goal" instead of "Our purpose"?

p.2597 l.17-l.21: please specify what type of temporal interpolation was used between the 1000yr reconstruction snapshots in ICE-5gV1.2, or what is the update interval time for the ice-sheet forcing?

p.2597, l.28 additional reference to Friedrich, T., A. Timmermann, L. Menviel, O. Elison Timm, A. Mouchet, and D.M. Roche, The mechanism behind internally generated centennial-to-millennial scale climate variability in an earth system model of intermediate complexity, Geosci. Model Dev., 3(2), 377-389, doi:10.5194/gmd-3-377-2010, 2010.

p.2598, l. 21: write "North Atlantic"

p.2598, l.26-28: please rewrite sentence. Unclear what is the message.

3 Analysis Method

p.2599 formula and text l. 19: please use σ^2 as symbol for the variance. This is the standard convention in many statistical textbooks (σ refers to standard deviation).

p.2599 significance test: the p-value of 1.962 is for two-sided test? If only tested for temperature increase, then significance is for a one-sided test (2.5%). Note: the significance is usually expressed for the probability of rejecting the null hypothesis when means are equal, i.e. 100%-95% = 5% significance level.

4.1 Results Annual mean

p.2600, l.6-16: Note that seasonal response and/or seasonal feedback can result in a precession signal in the annual mean.
p. 2601, l. 10: write: [...] between 30ºN and 30ºS. [...]
p.2601, l. 22: "series" instead of "serie"
p.2601, l. 23: [...] by those two [...]

4.2 Seasonal means

p.2602, l.5-7 : this section is introduced as a 'confirmation' of the annual mean results, but it is a more detailed view on the complex season/regional

timing of deglacial warming, since DJF and JJA behave quite differently.

p.2602,l.15: "northern North"

l.23: write: [...] linked to shrinking sea-ice extend in winter.

4.3 Precipitation evolution

p. 2603 l.1: write: tropical regions, the main [...]

l.3 [...] in annual precipitation.

l.3-11: Is this the correct interpretation:
    The test for differences in precipitation is a two-sided test. The significance
    level is now 5% for a two-sided test (i.e. you tested if abs(t) is > 1.962?

l.11: define acronym ITCZ here: "Intertropical Convergence Zone (ITCZ)"

l.12: suggested writing: "The annual mean precipitation shows a significant
    decrease in a zonal belt in the southern equatorial regions during 20 and 16
    kyrs BP."

l.14: write [...] of the ITCZ in response [...]

p.2604,l.10 : "northern Africa"

l.10-12: suggested citations: (1) deMenocal et al., Quat. Sci. Rev. , 2000; (2)
    Tjallingii et al., Nature Geosc., 2008; (3) Timm et al., J. Clim., 2010

4.4 Impact of interannual variability

p.2604 l. 14- p.2605 l.9: The text is okay, but maybe the authors could refer
    back to Section 3, the equation for the t-statistic. One could emphasize that
    two factors are defined by the climate system itself (the differences in
    means, and the internal variability), whereas one has the 'freedom' to
    adjust the sample size. The authors had to find a trade-off between the
    power of the test and the timing accuracy of the first significant change.

Robustness was tested by changing the sample size.
p.2605, l. 13: write: " In the northern tropical regions over the Pacific and southwestern North America [...]"

5 Discussion

p.2606, l. 10: suggested writing: "To reproduce the effects of millennial-scale climate variability, the modeling study would require the use [...]"
p.2606, l.16-l.22: The discussion of the use of ensemble simulations for detecting the timing of first changes: If the temporal samples could be replaced by ensemble samples, how large should the ensemble size be? 25, 100, 300 members?
The greatest advantage might be that one could test the changes at every year, instead of a window of 25,50,100,300 years. Therefore, the timing accuracy would be better.
p. 2606, following l.23: One should discuss also how the choice of the significance level affects the outcome of the test. More conservative significance levels reduce the type-one-error (rejecting the null-hypothesis [means are equal] when they are equal). On the other side, with a larger t-value one might risk to miss an early small change (type-two error: accepting the null hypothesis [means are equal] even though the samples have different means. Fig. 6 appears to me that type-2 error is likely encountered in the t-test at 5% significance.)

p.2607, l. 1-17: The implications from the model analysis for the proxies should be extended by at least two more questions of interest:
(1) Even if proxies were perfect records of past climate variability without dating uncertainty, there is a statistical limitation in the accuracy of determining the first deglaciation signal in single proxies.
(2) the model results could provide guidelines which geographical regions could provide proxies with the best signal-to-noise ratio for identifying externally driven climate signals, and how wide the sampling window should be.

Figures:

Fig.2: Caption: [...] anomaly to the 0-21 kyrs BP mean [...]

Please describe more precisely the winter/summer season (mid-month values for Jan/ Jun?

Fig.4: [...] from a 100-yr sample at 5% significance. Color scale is the date in kyrs BP. Black denotes [...]

Fig. 5: write "color scale"

Fig. 6: Optional: This really illustrates great the difficulties with objective statistical testing: If you could indicate the mean and standard deviations in the PDFs, this would be appreciated. Furthermore, is there a way to mark the 95% confidence range equivalent to the t-test in the red PDF? Then, one could use this illustration to refer to in the discussion of the statistical test procedure (i.e. Type-1 error and type-2 error).

Fig. 7-Fig.9: write "Color scale"