Review of the article entitled: “Mid-to-late Holocene Temperature Evolution and Atmospheric Dynamics over Europe in Regional Model Simulations” by E. Russo and U. Cubasch

Summary:

The manuscript presents a set of time slide simulations with a RCM for the mid-to-late Holocene. The skill of the simulations is stabilised comparing 2 m temperature and precipitation to other datasets for a reference period, and the output of the simulations for the past periods is compared to temperature reconstructions derived from pollen archives. Finally, potential sources for the disagreements identified and sought.

General comment

The paper can be broadly divided into three parts: i) validation of the simulation, ii) comparison with reconstructions iii) search for explanations of the disagreements. In all these three elements, I can identify caveats that in my opinion should be improved with different/complementary analyses. I have tried to review them in a comprehensive, yet constructive way, as detailed below.

Besides the technical aspects, I think there is room in the manuscript for improvement regarding writing style. It was challenging for me to read and understand many parts of the paper. This is in part due to incomplete information in the captions and the main text, wrong labelling in the figures, and the misleading use of some concepts as “observation” or “validation”. The internal structure in the paragraphs is confusing: paragraphs loosely connected, overly short, or in a misleading order respect to the panels in the figures. These issues add complexity that makes the lecture of the paper uncomfortable. Despite this rather negative view, I try to be constructive giving a list of points that develop the aspects that in my opinion can be improved in the manuscript. Note however that this list is not comprehensive.

Comments regarding the abstract/introduction

In the first line, “is always been mentioned” is grammatically wrong. Despite that, it sounds a bit loose, almost sceptical. Is it an important factor or not? This ambiguous tone of the first sentence of the abstract is manifest through the whole manuscript. By the way, the authors do not make an attempt to demonstrate that this is indeed the case for these simulations. More on this below.

The paper is somewhat optimistic regarding the use of “for the first time”. It is true that, as far as I know, there are no other set of time slide simulations. However there are various high-resolution simulations for the last millennium for Europe. Actually there exists at least one transient simulations for the last two millennia, in fact driven by the same ECHO-G run used by the authors of the manuscript. The authors should not ignore such previous, yet scarce efforts in this topic in the intro, but also the discussion of the results.

In line 6, “validation” is used in a wrong context. The model is validated normally against observations. But you can not validate the model looking at a reconstruction. Neither you validate a reconstruction looking at a simulation. You can only compare them, and try to gain insight through the disagreements. The use of “validation” in this wrong context is spread through the manuscript and should be avoided.

I think at least the first four paragraphs can be safely merged.

Lines 39 to 41 it is argued that then changes in solar irradiation were “negligible”, and latter than “we expect that such changes would imply relevant variations...”. It sound contradictory.

The paragraph in lines 42 to 45 is made out of a single sentence, which is too long. Still, the paragraph itself is short and can be merged with the former. Further, such sentence demands references.

In the paragraph starting in line 89, some examples of RCM simulations in palaeoclimate applications are outlined. It's strange to see that no simulation for Europe is referred. Examples of such simulations are: Gómez-Navarro et al. (2011, 2012, 2013, 2015a, 2015b) and Schimanke et al. (2012).

Comments regarding the model validation

It is not clear how long is the control period used in section 3.1. The only hint is the label in Figure 2, 1990-2000. Is that the case? It should be clearly stated, not only in the main text, but also in the caption of the figure. Actually, the length of this period is CRITICAL for the model evaluation, a fact that is not acknowledged in the discussion of the results. A 10-year period of a GCM simulation is strongly populated with internal variability. Under this scenario, a comparison with observations is tricky. The model could be "by chance" going through a cold or warm phase, which would have a strong impact in the validation, at least in the way it has been established in the paper, focused on mean values. In this sense, the validation does not look at important aspects such as the variability. How is the variance reproduced by the model? I'm not sure due to the short length of the simulation, but it could make sense to look at the variability modes of temperature and precipitation.

I do not think the choice of target for the validation is the best one. Using ERA Interim for the validation of precipitation in particular is a bad idea, since it is not constrained by precipitation observations, so there is no warranty that this dataset is bias free. I think it would be wiser to use the E-OBS dataset, which was developed specifically for the validation of RCMs in Europe (Haylock et al. 2008).

The way the similarity between the model and the observations is presented is a bit confusing to me. When the difference between two normally distributed variables is shown, the standard and intuitive approach, which steams from the application of the Central Limit Theorem, is to apply a t-test. The KS test is more suited for testing the shape of PDFs when the mean is know to be the same. For example if two dataset have the same mean, but different variance, the figure would show null bias (yellow colour here), but still the test would produce significant differences, which is misleading for the reader.

I think the maps showing precipitation difference are not very useful. A difference of 5 mm/day might be huge or tiny depending on the mean precipitation. I think changes in precipitation are more meaningful when shown as perceptual deviations with respect to the mean.

It is mentioned that the dots indicate grid where differences are "significantly not different". That's not exactly true. They indicate areas where the null hypothesis of the data being sampled from the same underlying distribution could not be ruled out, i.e. where they are "not significantly different".

I agree with the hypothesis used to explain the model deficiencies in Souther Europe regarding soil-atmosphere feedbacks. The particular role of these processes in RCM simulations in areas with strong water deficit was investigated in detail by Jerez et al. (2010, 2012).

Line 206 reads "These findings CONFIRM that… are MOST PROBABLY...". This is an example of doubtful and confusing sentence that should be avoided.

Maybe is worth to mention that generally the model skill resembles that identified in similar simulations for Europe (Schimanke et al. 2012, Gómez-Navarro et a. 2011, 2013).

Comments regarding the comparison with pollen reconstructions

As pointed out above, this comparison is by no means a model validation. This should be made clear in the wording. As such, all sentences like “CCLM performs well” should be modified.
The maps in Figure 5 are calculated as means with respect to which period?

In Figure 6, error bars are provided for the pollen data, but not for the simulation. I'm aware it is not easy to stablish them. However, such errors/uncertainties should not be neglected in the discussion of the results. The model has deficiencies that introduce systemic biases. But on top of then, there are non systematic biases introduced by unpredictable internal variability. This factor might lower or rise mean temperature in the simulation quite significantly, as pointed out by Gómez-Navarro et al. (2012) in a very similar scenario. Thus, this should be discussed at least qualitatively in this part of the text.

Many conclusions are drawn from Figure 6 regarding matchings of trends. I'm not sure at what extent such conclusions have any statistical significance, since in almost all cases the simulation lies within the uncertainty of the reconstruction. Having an almost perfect match between the reconstruction and the simulation is still perfectly possible within the range of uncertainty of the reconstructions.

Something I miss in this analysis here is the GCM simulations used to drive COSMO. I wonder how the ECHO-G and later ECHAM5 compares also with reconstructions. Is the RCM adding anything relevant to these simulations? If the answer is “certainly yes”, then the use of the RCM is fully justified and the paper would gain interest. If the answer is “mostly no”, it would be still interesting, since it would imply that the many GCM simulations available for the last millennia are still relevant at rather regional scales. I'm sure the PMIP community would be very interested in answering this question.

Comments regarding the interpretation of the paleo records

Generally it was difficult to follow the arguments in this section. It would significantly help to label the maps as Fig 6b, Fig 6c, etc. and use such labels extensively through the manuscript. In this regard, the discussion of the results starts with summer, whereas the first row shows winter. Small inconsitences like this, although non critical for the scientific message, have a dramatic impact in the reading pace.

The EOFs for MSLP are shown and used in the discussion. They are used to argue regarding NAO and SNAO, for instance. I'm not totally comfortable with that, since the NAO is defined as the leading pattern for a spatial window that is not that of the RCM. This explains in my opinion why the NAO pattern does not stand out as the leading mode in winter, and second mode in summer just “resembles” the SNAO pattern. I think a more orthodox approach would be to calculate the EOFs within the GCM, in a window that properly encompasses the North Atlantic. This is justified since the large scale circulation is fixed by the GCM, and thus the NAO simulated should be consistent with the climate variability within the RCM domain. Hence, such patterns could still be used to discuss about regional variability within the RCM domain.

Line 255 reads “In summer the first EOF shows that the model reproduces similar conditions in atmospheric circulation between the mid-Holocene and pre-industrial times”. I do not understand how that conclusion is drawn from the map in Figure 8.

In page 8 the wording “observed” is used in various sentences, and it's not fully clear what is meant (most likely respect to the simulation, but it could also be the reconstruction). I think “simulated” is more appropriate and precise.

Some inferences about the “clearness” of the sky are made which are based in indirect evidence such as EOF analysis. I think it is not necessary to make such risky affirmations. We have direct information that can tell us exactly how cloudy the simulated climate was. After all, in the simulation we can check directly variables such as cloud cover, which give a direct measure of what is being argued. I would go for a direct measure whenever possible, as it is the case. Similarly, in the paragraph between lines 277 and 279 (and Fig. 11) the more pronounced positive phase of the NAO can be directly tested within the GCMs, rather than indirectly inferred through a map of temperatures.

Finally, I think there are more powerful statistical tools than the one used here to study the co-variability between temperature and MSLP. Canonical Correlation Analysis could be used to derive relations between the variability of MSLP and temperature, and it would produce a picture of such co-variability more robust that the one provided by maps in Figure 10, for instance. An example of the application of such a tool in a very similar context is Gomez-Navarro et al. (2015b)

Comments regarding Figures

Figure 1: The colour scale shows everything below 1000 meters as green. I think a palette with stronger contrast could be chosen.

Figure 2: The reference period should be stated in the caption. I think the limits of the palette can be adjusted to better span the range of temperatures.

Figure 3: The colour palette provides barely any contrast all. Everything is yellow in the maps.

Figures 4 and 5: Same comments as in former figures

Figure 6: Please label panels as 6a, 6b, etc. I do not think using colour in the caption is an orthodox approach. Note that the caption does not agree with the order of panels. First row does not show North, but it is the first column which does, etc.

Figure 7: I can barely see the numbers and labels in the figures in the right.

Figure 8: I think the label with the loading can be moved to inside the maps. This would allow to put the maps closer together, which would allow to make maps larger and more readable. The latter comment can be applied to almost all figures.

Figure 9: Please label panels to indicate which represent EOF1 etc. Where are the units? Either the EOF or the PC carries the units, in this case pressure. I guess they are included in the EOF patterns in Figure 8. If so, please label the palette accordingly.

References

Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P., Lorente-Plazas, R., González-Rouco, J. F., and Zorita, E.: A regional climate simulation over the Iberian Peninsula for the last millennium, Clim. Past, 7, 451–472, doi:10.5194/cp-7-451-2011, 2011.

Gómez-Navarro, J. J., Montávez, J. P., Jiménez-Guerrero, P., Jerez, S., Lorente-Plazas, R., González-Rouco, J. F., and Zorita, E.: Internal and external variability in regional simulations of the Iberian Peninsula climate over the last millennium, Clim. Past, 8, 25–36, doi:10.5194/cp-8-25-2012, 2012.

Gómez-Navarro, J. J., Montávez, J. P., Wagner, S., and Zorita, E.: A regional climate palaeosimulation for Europe in the period 1500–1990 – Part 1: Model validation, Clim. Past, 9, 1667–1682, doi:10.5194/cp-9-1667-2013, 2013.

Gómez-Navarro, J. J., Werner, J., Wagner, S., Luterbacher, J., and Zorita, E.: Establishing the skill of climate field reconstruction techniques for precipitation with pseudoproxy experiments, Climate Dynamics, 1–19, doi:10.1007/s00382-014-2388-x, 2015a.

Gómez-Navarro, J. J., Bothe, O., Wagner, S., Zorita, E., Werner, J. P., Luterbacher, J., Raible, C. C., and Montávez, J. P: A regional climate palaeosimulation for Europe in the period 1500–1990 – Part 2: Shortcomings and strengths of models and reconstructions, Clim. Past, 11, 1077-1095, doi:10.5194/cp-11-1077-2015, 2015b.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, D20119, doi:10.1029/2008JD010201.

Jerez, S., J. P. Montavez, J. J. Gomez-Navarro, P. Jimenez-Guerrero, J. Jimenez, and J. F. Gonzalez-Rouco (2010), Temperature sensitivity to the land-surface model in MM5 climate simulations over the Iberian Peninsula, Meteorol. Z., 19(4), 363–374.

Jerez, S., J. P. Montavez, J. J. Gomez-Navarro, P. A. Jimenez, P. Jimenez-Guerrero, R. Lorente, and J. F. Gonzalez-Rouco (2012), The role of the land-surface model for climate change projections over the Iberian Peninsula, J. Geophys. Res., 117, D01109, doi:10.1029/2011JD016576.

Schimanke, S., Meier, H. E. M., Kjellström, E., Strandberg, G., and Hordoir, R.: The climate in the Baltic Sea region during the last millennium simulated with a regional climate model, Clim. Past, 8, 1419–1433, doi:10.5194/cp-8-1419-2012, 2012.