Review on

M. Forrest et al.

"Climate-vegetation modelling and fossil plant data suggest low atmospheric CO_2 in the late Miocene"

Climate of the Past Discussions, 11, 2239-2279, 2014

July 16, 2015

General comments

The paper presents the results from four simulations with the LPJ-GUESS dynamic global vegetation model (DGVM) driven with climate data for the Tortonian obtained from two AOGCM simulations using 280 and 450 ppm CO_2. The resulting global vegetation distributions are compared with proxy data from about 170 sites (mostly located in temperate regions), with results from similar simulation studies, and with additional evidences on Tortonian vegetation e.g. from fossil mammals or phytoliths.

Methodologically, the authors distinguish between an analysis at global scale (section 4.2) and an analysis at regional scale (section 4.3). While for the global analysis they introduce an "agreement index" to compare the site data with simulation data, the analysis at regional scale is almost completely qualitative. At both scales the authors conclude that paleo evidence is in better agreement with a lower CO_2 value. By their particular simulation setup, they also conclude that its mostly the climate effect of CO_2 that determines the resulting vegetation distribution and not the physiological effect of CO_2 fertilization.

There are only few studies of Tortonian climate taking advantage of the knowledge on vegetation-climate interactions encrypted in DGVMs. Insofar, the study provides a timely contribution to the research on pre-Quaternary climates. But methodologically the paper could be improved in three aspects:
First, the statistics behind the comparison between fossil data and model results is not really convincing. Partly this may be because the authors tried to keep the presentation short, but more fundamentally, important aspects of a robustness analysis of their statistical approach are missing (details follow below).
Second, the regional analysis (section 4.3) is rather unrelated to the global analysis (section 4.2), although it would be easy to repeat the statistical analysis performed globally also regionally. Surely, the data base is quite small for some continents, but by adding such an analysis one would get a clear impression why at a regional scale the study must stay qualitative.
Third, in the regional discussion a clear concept is missing for judging whether the differences seen in PFT distribution, biome distribution, tree fraction, and grass fraction between the 280 ppm and the 450 ppm simulation results are large enough to allow an

interpretation towards a higher or lower atmospheric CO_2 concentration. Therefore, I do not see that this qualitative discussion is appropriate to vote for or against a high or low CO_2. Instead, I would suggest to consider this qualitative regional analysis to be a check for the consistency of the continental vegetation patterns seen in their simulations with results from simulations of other groups and with evidences from aditional fossil data.

More detailed comments

1. Visual inspection suggests that the difference in biome distribution between simulated and reconstructed potential vegetation for today (Figs. S1A and S1B in the Supplement) is larger then the simulated Tortonian differences between low and high CO_2 (Figs. 1A and 1B). If this were true, the authors should explain why they can derive the main result of their paper from simulations that are within the range of model errors. I suggest that the authors apply a rigorous similarity/dissimilarity statistics to their biome distributions to quantify the model errors and compare them with the size of the signal they intend to interprete.

2. The concept of the "Agreement Index" needs further explanation. I failed to understand how the "fractions" that characterize PFT status are obtained from LPJ-GUESS. It is said that they are derived from the LAI (p. 2249, line 19), but the authors did not explain this relation.

3. In view of the various problems with paleo-botanical data, there is indeed no ideal way to compare them with model results. And surely the Agreement index (AI) introduced by the authors could be one way to quantify agreement. Nevertheless, this index is based on a number of arbitrary decisions: (i) the choice of fractional ranges for the different PFT 'statuses', (ii) the choice of numbers for the quantification of the different types of agreement (table 1); and (iii) the choice of the null hypothesis. To explain the latter a bit more: Instead of assuming that all possible values for the agreement (values -2 to 2) have equal probability, one could also assume that all fractional values for the "data" and the "model" have equal probability which would give a different random distribution ("null" distribution) of AI values.
In my opinion there is no good argument for either of the choices (i) to (iii). Therefore it is not clear whether the results based on the particular choices for the AI are robust. The authors claim to have addressed robustness with respect to (i), but did not present these results. Robustness with respect to all aspects should be demonstrated in the paper (or in appendices) by varying the particular assumptions (i) to (iii).

4. The arguments for introducing the new AI measure of data-model agreement (p. 2249, lines 13-17) are not convincing: The authors simply state a personal preference ("We prefer a metric that ...") but do not explain why the other metrics (Salzmann et al. 2008; Pound et al. 2011; Francois et al. 2011) should be discarded. In fact, it would be good to know whether those other approaches would reveal similar results when applied to the data used by the authors. I personally feel, that in particular the method by Francois et al. (2011) is the most objective because it generally distrusts a comparison of data diversity with model abundances (in the terminology of the authors, p. 2248 bottom) by comparing only presence/absence. Moreover, if dispite all warnings such a diversity-abundance comparison is attempted (as done by the authors with their Agreement Index), why not using the classical rank correlation which is known to be statistically robust?

5. With Fig. 2 the authors want to demonstrate that their results differ from the null hypothesis of random agreement. And indeed, the AI values for the 280 ppm and the 450 ppm simulation are well off their "null model". But they did not demonstrate that the difference between the AI values obtained from their two simulations with different CO_2 is significant. If naively one would add the spread of the null model to the AI values from the two simulations, they would be statistically indistinguishable. Therefore the authors must plot into Fig. 2 also the full distribution of their results for the two experiments to allow judgement of signifance concerning their difference – maybe the authors added those Z-scores exactly for that purpose, but it's nor how they were computed. But plotting the individual distributions would in any case be more informative.

Minor comments

p. 2246, line 25: The authors note that they transfered the soil parameters of the AOGCM to LPJ-GUESS. This provokes the general question to what extent the water cycles in the AOGCM and LPJ-GUESS are consistent, and whether inconsistencies in evapotranspiration fluxes might affect the results for the vegetation distribution.

p. 2247, lines 18-28: The authors describe a number of modifications they introduced to LPJ-GUESS, but not why these modifications were necessary for their study. For the modified bioclimatic limits they claim improvements for present day biome distribution (lines 18-20) but do not demonstrate the improvements. It is only claimed (p. 2248, lines 11-12) that the modern biomes are reproduced "reasonable well". For such a claim one needs a measure, but this is not provided. Moreover, the main issue of the study depends on the model's reaction to changing climate and CO_2. Therefore, some comments why the authors trust the model's response to such changes would be helpful.

p. 2251, lines 10-11: Here the authors announce a table in the supplement relating fossil plant taxa and PFTs. But such a table is missing. Please add that table since a large part of the study is based on this classification. Instead there is an un-numbered table in the supplement listing the study sites.

p. 2252, line 16 and Figs. 1a an d 1b: It would be good to refer to Appendix B for references to the biome classification. Even better in my opinion would be to serve the readers by providing a table with the rules for the biome classification.

p. 2255, line 7: What are the "two reasons"? I cannot identify them in the following text.

Table 1: I guess the row headers should be shifted.

Supplement Fig. S2: This figure should in my opinion be shifted to the main part of the study, because it shows that in certain regions (e.g. the Iberian peininsula) the proxy-data are not informative about the value of atmospheric CO_2.