Climate-vegetation modelling and fossil plant data suggest low atmospheric CO_2 in the late Miocene (cp-2015-65)

M. Forrest, J Eronen et al.

We thank the reviewers for their insightful comments which have improved the manuscript considerably. Below we have answered their comments, and provided further information and data concerning how we have integrated their suggestions in the revised manuscript. A list of changes and the revised manuscript with changes marked follows the response to reviewers in this document. The reviewer comments are in black, our answers are in light blue.

Reviewer #1

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 CO2. 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 CO2 value. By their particular simulation setup, they also conclude that its mostly the climate effect of CO2 that determines the resulting vegetation distribution and not the physiological effect of CO2 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).

We had previously performed multiple robustness tests for the analysis but as the reviewer mentions, most of these were left out of the original manuscript because we wanted to keep the presentation short. We agree that these explore an important aspect of the novel method presented here. We have provided these robustness checks and addressed all of the more detailed points raised by

the reviewer below in supplementary information to the revised manuscript as requested.

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.

We think that the regional analyses and discussion of these are important and particularly interesting for researchers with a regional focus. We fully agree that applying the statistics at the regional scale might not be very meaningful, not only because of the small sample size (e.g. only three sites in Africa), but also because we cannot expect a global vegetation model driven by a global climate model to be very accurate at the regional scale. To illustrate the limited coverage of the fossil database, we have combined Table 2 and Table 3 of the original manuscript and included AI scores from all continents as well as the number of fossil sites in each region. The Central Europe region was enlarged compared to the previous version to include more data points; this does not affect the conclusions. Furthermore, for the discussion of regional scale aspects, we also rely on other independent evidence, such as fossil mammals, phytoliths and isotopes that indicate open conditions for North America.

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 CO2 concentration. Therefore, I do not see that this qualitative discussion is appropriate to vote for or against a high or low CO2. 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.

We thank the reviewer for raising this point. We agree that we might have stretched the regional interpretation in the manuscript. We have corrected this in the revised version, focusing more on evaluation compared to other studies and discussing the differences between the 280ppm and 450 ppm scenarios, but only mention an indication of lower or higher CO2 concentrations when the pattern is very clear, such as in North America, where the more open vegetation under low CO2 clearly corresponds better with the paleobotanical data and other independent sources of evidence. We now focus more on how well our model produces the regional and continental vegetation patterns during the Miocene (as compared to paleobotanical evidence and other modelling studies). The proxy data include well-known samples from fossil mammals, isotope data and sedimentary records from Europe and North America.

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 CO2 (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.

We agree that in its original form the manuscript does not present sufficient analysis of the model uncertainties and signal size. We were reluctant to use the statistical similarity/dissimilarity metrics to analyse biomes for our main comparison for reasons that we outline below. However, we agree with the reviewer that a statistical comparison can provide useful insights. Therefore, we have now evaluated our simulations with Cohen's Kappa statistic, which is a standard for comparing modelled biomes. The results show acceptable agreement between our present day simulation and the reconstructed potential natural vegetation. We have also used Kappa to quantify the difference between the modelled biomes and find that our model setup can distinguish the two Tortonian scenarios from each other and from the present day control run. The results are detailed below and in the supplementary material accompanying the revised version of the manuscript.

Drawbacks of using Cohen's Kappa for biome comparisons

The first drawback of comparing Kappa scores for biomes is that Kappa does not include any "degree of difference" mechanism which can be important when considering more than two categories. For example, there is a much smaller conceptual difference between a "tropical grassland" and a "tropical savanna" than there is between a "tropical grassland" and a "boreal evergreen forest", but that difference is treated identically when calculating Cohen's Kappa. This can be ameliorated to some extent by aggregating to megabiomes as done by Pound et al. (2011) (an approach we now follow), but is inevitably present to some extent. A weighting can also be attempted, but this introduces subjective decisions.

The second argument against comparing potential natural vegetation (PNV) biome distributions using Kappa is that PNV biome classifications themselves introduce uncertainty. Potential natural vegetation cannot be measured directly (it no longer exists due to human influence) and so must be reconstructed. There is uncertainty in such reconstructions as evidenced by the differences between PNV biome maps: for example, the horn of Africa is predominantly covered by "tropical deciduous forest" in Haxeltine and Prentice (1996), but is dominated by "dense shrublands" in Ramankutty and Foley (1999). Similarly, the extent of the "tropical deciduous forest" biome in Southern Africa varies considerably between the two maps. Even the biomes categories themselves vary between the maps as different authors make different distinctions. Our experience is that kappa statistics applied to compare different PNV maps can

indicate as bad agreement as the one between a model and a PNV reconstruction, when biomes are not aggregated to coarser classes. There are also subjective choices when classifying model output which introduces uncertainty. For example, how much tree LAI or tree cover constitutes a forest? How much for a savanna? The choices for these numbers are not well-motivated and can change the biome boundaries considerably. Concerning the paleobotanical data, we deliberately did not derive biomes because classifying fossil sites into biomes introduces large uncertainty arising from interpreting the fossil record in terms of vegetation cover.

These arguments are now included in Section 3.4.1 of the revised manuscript.

Quantifying Model Uncertainty using Kappa

We have compared our present day control run with a reconstructed biome distribution (e.g. Hickler et al. 2006) using Cohen's Kappa. To mitigate the sources of uncertainty listed above, the data were aggregated to megiobiomes following the approach of Harrison and Prentice (2003) and Pound et al. (2011). The results show acceptable agreement between our present-day simulation and the PNV reconstruction, with a Kappa score of 0.62, constituting "good" agreement by Monserud and Leemans (1992) However, the pure numbers should not be over-interpreted for the reasons we outlined above. This result and method are described in Section S3 of the supplementary material accompanying the revised version of the manuscript. We have also included a mention that a more detailed examination of the biomes produced by LPJ-GUESS (without the modifications for this study) has been done by Smith et al. (2014, their Figure 2(C))

Quantifying Effect Size using Kappa

Comparing the megabiome distribution from 280ppm and 450ppm Tortonian runs gives a Kappa of 0.70. Given that these biome maps are produced with identical methodologies (they use the same model structure differing only by the effect of CO_2 concentration on vegetation and climate, they utilise the same biome classification and hence have the same subjective choices, and they involve no data-originating uncertainty), we argue that we do see a sufficiently large signal for our interpretations.

Furthermore, the Kappa between the Tortonian 280ppm megabiomes and the PGF control run megabiomes is 0.64. Considering again that these maps are produced with identical methodologies, this indicates that we can distinguish Tortonian vegetation with 280ppm CO_2 and present day vegetation (in answer to reviewer 2's second point). Comparing the Tortonian 450ppm megabiomes and the PGF control run megabiomes gives a Kappa of 0.48. These scores are included in Section S3 of the revised supplementary material.

In summary, we believe that our vegetation model uncertainties are reasonable (given the uncertainty in the method of quantification) and our effect sizes are large enough to support our interpretation. We have included this information in the supplementary material. Note also that we used a DGVM that has been generally benchmarked and used for climate impact studies in a very large number of studies (see http://iis4.nateko.lu.se/lpj-guess/LPJ-GUESS_bibliography.pdf for a list of LPJ-GUESS publications)

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.

We have included further elaboration of the method in the manuscript. To answer the reviewer briefly here: the "fraction" (or "relative abundance") of a PFT in a gridcell is the LAI of the PFT in the gridcell divided by the total LAI in the gridcell. The LAI values are the growing season maximum values and they are averaged over a 30 simulation year period.

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).

Yes, we agree with the reviewer that we should have provided more information about the robustness of the method. We have addressed all of the above points and included our findings in Section S2 of the revised supplementary material. We have also significantly reworked the text discussing the quantification of agreement by chance. We realise that this text was too brief and did not clearly and fully describe the method, nor did it describe the possible choices or rationale for our choice. In particular, from this comment and comment 5, it appears that the reviewer misunderstood our method for estimating agreement by chance. We hope we have corrected this failure of the text and can also add

that, somewhat fortuitously, the reviewer also suggested two alternative methods which we have now tested and discuss in the supplementary information.

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; François 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 diversityabundance 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?

We thank the reviewer for pointing this out and agree we should be more exact in our reasons for developing the AI rather than using the other methods. We have included a more detailed discussion of the reasoning for not using existing methods or classical statistics in a revised draft of the manuscript and also present them below. As mentioned above, we will also provide additional statistical analyses of the AI method to prove the robustness of our results.

Furthermore, we have calculated both Pearson's product moment correlation coefficients and Spearman's rank correlation coefficients for the 280ppm and 450 ppm scenarios per PFT and for the entire dataset. These are now presented in the revised supplement to the main text (Section S1) and also summarised here for convenience in Fig 2. As mentioned in the original text, these do not prove to be particularly illuminating. The per-PFT coefficients do not show a consistent trend favouring a particular CO_2 scenario. Furthermore, the Spearman's rank for the full dataset is virtually identical for both CO_2 scenarios, but the Pearson's coefficient indicates better correlation for the 280 ppm CO_2 scenario than for 450 ppm CO_2 (0.53 vs. 0.42). This could be interpreted as weak evidence that the 280 ppm CO_2 scenario agrees better with the paleo-botanical data. We have included a brief discussion of these additional analyses in the supplementary material, and as indeed not all applied statistics clearly favor the low CO2 scenario, we will emphasize the uncertainties more. Note that we already formulated the title quite carefully, as: "Climate–vegetation modelling and fossil plant data suggest low atmospheric CO2 in the late Miocene." The wording "suggest" should indicate that we cannot be sure, as often the case in paleoclimate research. However, one should keep in mind that our qualitative regional discussion (where supported by sufficient data) also tends to favor the low CO2 scenario.

Regarding the other comparison methods; Salzmann et al. (2008) present a map of the inconsistency between model and data. Whilst a visual comparison is

useful, we wanted to add a quantitative method to discriminate between the two CO_2 concentrations. The later study of Pound et al. (2011) uses Cohen's Kappa to determine biome agreement, both the 27 'native' biomes from BIOME4 and a 7 "megabiome" classification. This does offer a single statistic which could be used for hypothesis testing. However, (as discussed extensively in point 1.) there are drawbacks to using Kappa to compare biome classifications and with biome classifications themselves. So whilst comparisons of biomes are clearly useful visual aids and can be a useful cross-check (see our response to point 1), we decided to use only information on PFT fractions for our main analysis and therefore minimize subjective choices and classifications.

As the reviewer points out, the work of François et al. (2011) offers a method for determining agreement between paleobotanical data and simulated vegetation which percentage agreement per PFT based on presence/absence. These per-PFT scores could conceivably be combined to produce overall agreement scores, taking care that PFTs which are mostly absent from the fossil record do not unduly affect the final result. However, our study is different in nature to that of François et al. The study of François et al. was a regional study with a relatively high degree of taxonomic precision (i.e. a more detailed PFT set), whereas our study is global with appropriately coarser taxonomic resolution (i.e. a relatively simpler global PFT set). By means of example, there are 8 purely temperate PFTs in the CARAIB version used in François et al. 2011 compared to only 2 in the default LPJ-GUESS configuration and 4 in the configuration used in our study. Thus by exploiting a high degree of taxonomic precision, presence/absence data were used effectively in the regional study of François et al. In our global study, each PFT spans a much larger geographical extent and there are fewer PFTs at each site for which to make presence/absence comparison. Thus we expect the effective differentiating power of such presence/absence to be lesser. So rather than using detailed taxonomic resolution and presence/absence information, we sought to exploit the abundance/diversity fractions which we believe has useful information and so is worth attempting despite our previous warnings. For this reason we developed the Agreement Index and introduced statuses beyond presence/absence.

The Agreement Index also allows easy assignment of a zero-weighting when PFTs are absent from a site in both the fossil record and model (contribution in this case is zero). It also allows an (admittedly subjective) method to tackle the "degree of difference" effect which causes problems for Kappa analyses which involve more than two classifications with differing conceptual degrees of similarity, as mentioned in point 1. This is done by assigning the value -2 for very strong disagreement and the value +2 for correctly matching dominant PFTs, as this must necessarily include at least 50% of the PFT and defines predominant biome functioning. A similar effect could be achieved by weighting the Kappa scores depending on the degree of difference, but this would also require subjective choices. The subjective choices involved in this method are motivated in an obvious and transparent way and can be (and were) tested relatively easily (see point 3).

We have modified the text in the manuscript to explain the above arguments in more detail.

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 CO2 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.

We agree that we could have provided more information on the difference between the AI values from different models. It also appears that the text which explains the distribution in Fig 2 in the manuscript is unclear and we have attempted to remedy this. To clarify here, each of the 25,000 frequency counts in Fig. 2 is the mean AI score from matching all 167 fossil sites to 167 random gridcells (not counts of the AI per site or AI per PFT). Thus there is no meaningful "full distribution" to plot on Fig. 2 for the two experiments because each experiment only yields a single frequency count of the type plotted in Fig 2 (ie. the mean of all the 167 fossil sites compared to simulated vegetation). It may be that the "full distribution" to which the reviewer is referring is the 'per site' or 'per PFT' AI values (or 'per site per PFT' AI values) but that quantifies a different variability from that in Fig 2. The variability in AI between sites is not inconsiderable (see Figure 1 in the original manuscript for an idea of the variability between sites) but we don't believe this sheds any light on the issue of distinguishing the mean AI values of the two CO_2 scenarios. Similar arguments apply for the distribution of AI per PFT.

In the first instance, the distribution in Figure 2 shows the mean value of chance agreement. This seems to be clear enough, although we should add that this is only one particular method of estimating chance agreement. Many other methods are conceivable and a selection have been tested and are now reported in the supplementary information to the revised manuscript in answer to the reviewer's point 3.(iii). One can then look at the AI values for each Tortonian scenario and conclude that both scenarios do indeed offer better agreement than chance. In the second instance, the standard deviation of the same distribution aims to quantify the natural variability in chance agreement and so give and idea of how much better the Tortonian scenarios are than random chance, and how much better one scenario is than the other. The traditional p-value interpretation is relative to the model used to estimate chance agreement. In the case of the method presented in the main text, this would be the probability of getting a random combination of gridcells giving better agreement than the Tortonian scenario. These are $p < 10^{-8}$ and $p < 10^{-13}$ for the 450 ppm scenario and the 280 ppm scenario respectively. We can conclude, reassuringly but not surprisingly,

that both our reconstructions are very much better than chance. Furthermore, the 280 ppm scenario is clearly better than the 450 ppm but differences in such very small *p*-values are not helpful, so instead we report the difference in units of standard deviation (*Z* scores), in this case 1.7. We believe this difference sufficiently supports our conclusion that the 280 ppm run agrees better with the fossil record than the 450 ppm run.

We realise that this logic relies on the assumption that matching random model gridcells to the fossil record gives an adequate representation of chance agreement. We chose this method to present in the main text because it will give ecologically consistent PFT compositions (no unrealistic combinations of boreal and tropical PFTs for example) and so is a more useful test than some random numbers (which could give such unrealistic combinations). However, in the supplementary information to the revised manuscript we test other models of random chance and, with one exception, all other methods of estimating agreement by chance indicate that the 280 ppm simulation is better than chance agreement by at least 3 standard deviations (*Z*-score >3) the 450 ppm scenario is better by around 1.5 standard deviations, but generally much higher.

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.

Yes, in this model set-up each model has an independent hydrological cycle with different process representations, with the hydrological cycle of LPJ-GUESS being driven (in terms of input precipitation and temperature) by the climate from ECHAM5/MPIOM. It is certainly true that the evaporative fluxes will not be identical between the models, the different land surface properties and different process representations will guarantee that. However, the hydrological cycle of LPJ-GUESS is still fully internally consistent and has been benchmarked (as implemented in the related model LPJ-DGVM) in Gerten et al. 2004 and, for newer version of the LPJ-GUESS model, including the one we applied, by the LPJ-GUESS consortium (unpublished). Given the wide-ranging applications of LPJ-GUESS and LPJ-DGVM, we are confident that the representation of the hydrological cycle, including the evapotranspiration fluxes, to be sufficiently well-modelled to reproduce the broad patterns of Tortonian vegetation at this relatively coarse global scale. The study of the different hydrological in different models is an interesting topic in itself, especially the hydrological cycles in the models have been designed with very different aims in mind, but beyond the scope of the current work. Double simulations of the hydrology are inherent in each application of a DGVM driven by a GCM and cannot be avoided. The alternative approach would have been to use an existing land surface (and vegetation model) fully embedded within a GCM, but the land surface schemes of

GCMs do commonly used more simplified representations of the vegetation and simulated vegetation patterns have not been evaluated as extensively as for the LPJ-GUESS model. As this is standard procedure, we don't believe it is necessary to mention in the main text.

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 CO2. Therefore, some comments why the authors trust the model's response to such changes would be helpful.

With regard to the bioclimatic limits, the main effect was to remove treeless areas in South China, Argentina and Florida (see Smith et al. 2014, Figure 2(C) for the model version which does not include nitrogen limitation). This was an artifact whereby in these areas it was too warm for temperate trees to establish, but too cold for tropical trees, which resulted in treeless belts. In other words, there was a mistake in the model, which we corrected, with the main result that the model correctly simulates forests in south-eastern Asia. The other changes to bioclimactic limits were made for consistency with Sitch et al. (2003) and make very little difference. The introduction of Temperate Needle-leaved Evergreen (TeNE) trees, and the splitting of shade-Intolerant boreal/temperate Broadleaved Summergreen trees (IBS) into Temperate shade-Intolerant Broadleaved Summergreen trees (TeIBS) and Boreal shade-Intolerant Broadleaved Summergreen (BIBS) was intended to better compare the model results to the fossil record and because we believe that, with these changes, functional characteristics of the global vegetation are represented more appropriately. We have now described the reasoning for these changes in more detail in the revised text. With regards to the model's ability to capture present day biomes, we refer to our answer to point 1 which includes a Kappa measure and higher resolution maps for a more detailed visual comparison. We have also mentioned in the supplementary material (section C3 where model evaluation is discussed) that the biomes produced by LPJ-GUESS without our modifications can be seen in Smith et al. (2014) (their Figure 2(C)).

Furthermore, we have included text to mention that LPJ-GUESS (and the closely related LPJ-DGVM model) has been benchmarked against various observations including, for example, NPP (e.g. Zaehle et al., 2005; Hickler et al., 2006), modelled PNV (Hickler et al. 2006; Smith et al. 2014), stand-scale and continental-scale evapotranspiration (AET) and runoff (Gerten et al., 2004), vegetation greening trends in high northern latitudes (Lucht et al., 2002) and the African Sahel (Hickler et al., 2005), stand-scale leaf area index (LAI) and gross primary productivity (GPP; Arneth et al., 2007), forest stand structure and development (Smith et al., 2001, 2014; Hickler et al., 2004), global net ecosystem exchange (NEE) variability (Ahlström et al. 2012, 2015) and CO_2 fertilisation

experiments (e.g. Hickler et al. 2008; Zaehle et al. 2014; Medlyn et al. 2015). Many of these benchmarks are constantly repeated by the LPJ-GUESS consortium (of which Hickler is a member, unpublished). Regarding the CO2 response, the model without nitrogen limitation most likely overestimates CO2 fertilisation (see e.g. Hickler et al. 2015), which implies that our conclusion that the climate forcing is more important than the physiological CO2 effects for distinguishing the low and high CO2 scenario for the late Miocene is robust. This is now also discussed in the revised manuscript.

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.

This table has been added to the supplementary material.

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.

Yes, we have now included such a table

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

The two reasons are increased seasonality in Central Europe, and increased openness in the Iberian Peninsula and in modern Turkey. However, we agree that this is unclearly worded and this has been re-worded in the revised version of the manuscript.

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

Thank you for pointing this out, we will ensure this is correct in the final proofs.

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 CO2.

Yes, this is a good idea and we have done so.

Reviewer #2

This paper presents a reconstruction of late Miocene vegetation using a dynamic vegetation model driven by the climatic outputs of climate model runs for two different partial pressures of CO2 in the atmosphere, 280 and 450 ppmv. These partial pressures reflect the range of atmospheric CO2 pressures that have been reconstructed from proxy data for the late Miocene. The authors compare the vegetation reconstructed with palaeovegetation data available for this time period. They also compare in detail their results with late Miocene vegetation

model reconstructions published in the literature. For the comparison with the data, they build an agreement index (AI) which is an interesting and relatively novel aspect of their work. Since the AI is significantly higher for the low CO2 (280 ppmv) case, they conclude that climate and vegetation modeling suggest low CO2 in the late Miocene and so would favour the lower values in the range exhibited by the proxies.

The paper is generally well written, scientifically sound and with some clearly novel aspects with respect to previous work on the subject. I am thus in favour of its publication in Climate of the Past. I just have a few comments or suggestions that the authors might want to address.

(1) Section 3.4 : your comparison at the PFT level and associated statistics is presented as a new method for model-data comparison. However, as mentioned by the authors, François et al. (2011) have also performed a similar comparison at the PFT level, and contrary to what is said here, they also used the PFT diversity from the data (see for instance their table 7 and the comparison with model NPPs in their figure 6), although only presence-absence is used in their kappa calculation. What is the advantage of your AI index compared to the more traditional kappa method ? Kappa can also be averaged over sites or over PFTs. The statistical study on kappa presented here for AI (which is really interesting and the most novel contribution of this paper) is also possible for kappa. You just define more classes (abundance classes) that may also be involved in the kappa method, but actually have not been involved because of the large uncertainties on model PFT abundances. Models are certainly more robust in evaluating presence/absence than abundance. Morever, as mentioned in your section 3.4, it is not obvious that PFT diversity from the data can directly be compared to model abundances. Even presence/absence in the data may be uncertain due to the PFT assignment scheme in the data (see again François et al., 2011). This may also critically depend on the number of PFTs in the classification used. This might be discussed somewhat more, because the associated uncertainty might have some impacts on the conclusions reached.

We thank the reviewer for his insightful and positive comments. We apologise for mis-representing the work in François et al. (2011), we meant to state that PFT diversity was not used to provide a quantitative measure of agreement, and have amended the text accordingly.

Our reasons for not using Kappa and for using abundance data beyond presence/absence are detailed in our answer to reviewer 1's comment 1. We would also argue that the coarser taxonomic resolution of our global PFT gives sufficient robustness in terms of presence/absence and abundance to use abundance fractions. Furthermore, we agree that whilst it could be possible to use Kappa on model abundances classes (neatly avoiding the uncertainties of biome classification whilst still utilising abundance/diversity data); such a method would still suffer from the "degree of difference problem" where a mismatch between the absent category and trace category would be treated as

585 severely as a mismatch between absent and dominant categories. It also offers no
586 obvious way to remove or zero-weight the contribution from PFTs which are
587 absent in both the data and model at a given site. We have discussed these points
588 in the revised text.
589
590
591 (2) Section 4.1, figure 2 : it might be interesting to add on figure 2 the AI that
592 would be obtained with present-day (control run) model vegetation (when
593 comparing to palaeodata). Is it significantly different from the AI for the 450 and
594 280 ppmv late Miocene configurations ? If it is close to the 280 ppmv late
595 Miocene case, it might mean that your model is not fine enough to discriminate
596 between the present-day vegetation and the late Miocene one.
597
598 As described in our answer to reviewer 1's point 1, we have now provided
599 statistics to quantify the differences in modelled vegetation between today and
600 the Tortonian. The Kappa between the present day control run and the Tortonian
601 280 ppm run is 0.64 and the Kappa between the present day control run and the
602 Tortonian 450 ppm run is 0.48. Given that identical methodologies were used to
603 derive these biomes (i.e. using the same model), we argue that we our model is
604 indeed fine enough to discriminate. However, we don't think that presenting the
605 AI for the present-day vegetation is meaningful for addressing the research
606 questions addressed here.
607
608
609 (3) Section 4.3.1 : the characteristics of Miocene vegetation in Europe is indeed
610 as discussed here the widespread presence of temperate deciduous trees, with
611 some temperate evergreens in the south. Evergreens are however different from
612 present-day Mediterranean (drought-tolerant) evergreen trees, since data show
613 the presence (not dominance) of temperate evergreen perhumid trees. This is a
614 very important climatic constraint from the point of view of the data, while your
615 model does not separate between drought-tolerant and perhumid temperate
616 evergreen trees. The impact of this simplification on the results should be
617 discussed, or at least it should be mentioned. Also, your figure S2 indicates that
618 the SI index strongly varies from one site to the next. This is an important result
619 that shows that there are still some features that are not well captured by the
620 model (or possibly it might be a problem in the interpretation of the data). It
621 would be interesting to discuss figure S2 in the main text.
622
623 It is right that both evergreen types are lumped in the applied version. However,
624 the temperate evergreen PFT in this model version represents rather the
625 perhumid type. The special hydraulic features of the drought-tolerant type (e.g.
626 sclerophyllous leaves having a lower wilting point) had only been implemented
627 in one particular model version and application including the hydraulic
628 architecture of different PFTs, which improved the simulations for present-day
629 Mediterranean ecosystems (Hickler et al. 2006). These developments have, to
630 date, not been transferred to newer versions of LPJ-GUESS, partly because, back
631 in 2006, the computational demand was still limiting, and calculating all
632 physiological processes for each cohort would have increased the computational
633 demand by an order of magnitude. Now, the computational demand is not so

much limiting anymore, and we envisage including tree hydraulics also in newer versions of LPJ-GUESS, but this has not been done. We added a couple of sentences to discuss this, but we think that going more into details would be beyond the scope of this paper.

Concerning the site-to-site variation of the AI in what was formerly Fig. S2 (now Fig. 5 in the revised manuscript), much of this is related to the fossil data rather than the model output as the variation often occurs within one simulated grid cell. For brevity, we choose not to discuss the details variation or possible nuances in the fossil data as this is primarily a discussion of modelling results at a global scale, and the manuscript is already rather long. However in accordance with the wishes of reviewer one, we have moved the figure to the main part of the manuscript so this variation will be more readily apparent to the reader.

(4) Section 5 (Summary and conclusions): In view of the large uncertainties on climate models (including other boundary conditions than CO2), vegetation models and PFT classification, I am not sure that models can really provide a strong constraint on palaeo-CO2. It is interesting to learn that you model is more consistent with low CO2 in the late Miocene, but this is a very indirect constraint. I would suggest that you reformulate the last sentence of your conclusion to make the statement less direct (there are uncertainties and it may be model-dependent, so we may need to study the same problem with other climate/vegetation models).

We fully agree with the reviewer that there are still large uncertainties in climate models, the applied vegetation model and the applied analyses. We have been aware of these uncertainties, but apparently some of the formulations indicated too much certainty. Thus, we have reformulated the last sentence of the conclusions and other key sentences throughout the manuscript. We nevertheless believe that our indirect evaluation of two plausible CO2 concentrations for the Tortonian and other aspects of the manuscript (e.g. state-of-art climate modelling and DGVM applied to simulate Tortonian vegetation, novel approach for comparison with paleobotanical data, separating direct climatic and physiological CO2 forcing) represent an interesting contribution to the science on Tortonian climate and ecosystem dynamics.

(5) Some small typos:
P 2254, line 10: 'possibly because' P 2262, line 25: 'Fig 1a and b'does not correspond to the present-day biome map, it should be figure S1 P 2263, line 7: 'It ´ also shows a band P 2263, line 12: 'particularly shrubs'

Thanks for pointing these out, these have been corrected.

References

Ahlström, A., Miller, P.A. & Smith, B. 2012. Too early to infer a global NPP decline since 2000. Geophysical Research Letters 39, L15403.

Ahlström, A., Raupach, M.R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J.P., Friedlingstein, P., Jain, A.K., Kato, E., Poulter, B., Sitch, S., Stocker, B.D., Viovy, N., Wang, Y.-P., Wiltshire, A., Zaehle, S. & Zeng, N. 2015. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science 348: 895-899.

Arneth, A., Miller, P.A., Scholze, M., Hickler, T., Schurgers, G., Smith, B. & Prentice, I.C. 2007. CO2 inhibition of global terrestrial isoprene emissions: Potential implications for atmospheric chemistry. Geophysical Research Letters 34: L18813.

François L, Utescher T, Favre E, Henrot AJ, Warnant P, Micheels, A., Erdei, B., Suc, J.P, Cheddadi, R. and Mosbrugger, V.: Modelling Late Miocene vegetation in Europe: Results of the CARAIB model and comparison with palaeovegetation data. Palaeogeogr., Palaeoclim., Palaeoecol., 304, 359–378, 2011.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance – hydrological evaluation of a dynamic global vegetation model. Journal of Hydrology, 286, 249–270, 2004

Harrison, S. P., & Prentice, C. I. (2003). Climate and CO2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. Global Change Biology, 9(7), 983-1004.

Haxeltine, Alex, and I. Colin Prentice. "BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types." Global Biogeochemical Cycles 10.4 (1996): 693-709.

Hickler, T., Smith, B., Sykes, M. T., Davis, M. B., Sugita, S. and Walker, K.: Using a generalized vegetation model to simulate vegetation dynamics in northeastern USA. Ecology, 85, 519-530, 2004.

Hickler, T., Eklundh, L., Seaquist, J., Smith, B., Ardö, J., Olsson, L., Sykes, M.T. & Sjöström, M. 2005. Precipitation controls Sahel greening trend. Geophysical Research Letters 32: L21415.

Hickler, T., Prentice, I. C., Smith, B., Sykes, M. T. and Zaehle, S.: Implementing plant hydraulic architecture within the LPJ Dynamic Global Vegetation Model. Global Ecology and Biogeography, 15, 567-577, 2006.

Hickler, T., Smith, B., Prentice I.C., Mjöfors, K., Miller, P., Arneth, A. & Sykes, M.T. 2008. CO2 fertilization in temperate forest FACE experiments not representative of boreal and tropical forests. Global Change Biology 14: 1.12.

Hickler, T., Vohland, K., Feehan, J., Miller, P. A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T.R., Cramer, W., Kühn, I., and Sykes, M. T.: Projecting the

future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. Global Ecology and Biogeography, 21, 50-63, 2012.

Hickler, T., Rammig, A. & Werner, C. 2015. Modelling CO2 impacts on forest productivity. Current Forestry Reports 1: 69-80.

Lucht, Wolfgang, et al. "Climatic control of the high-latitude vegetation greening trend and Pinatubo effect." Science 296.5573 (2002): 1687-1689.

Medlyn, B.E., Zaehle, S., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hanson, P.J., Hickler, T., Jain, A.K., Luo, Y., Parton, W., Prentice, I.C., Thornton, P.E., Wang, S., Wang, Y.-P., Weng, E., Iversen, C.M., McCarthy, H.R., Warren, J.M., Oren, R. & Norby, R.J. 2015. Using ecosystem experiments to improve vegetation models. Nature Climate Change 5: 528-534.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. Ecological modelling, 62(4), 275-293.

Pound, M.J., Haywood, A.M., Salzmann, U., Riding, J.B., Lunt, D.J., Hunter, S. A: Tortonian (Late Miocene, 11.61–7.25 Ma) global vegetation reconstruction, Palaeogeogr., Palaeoclim., Palaeoecol., 300, 29-45, 2011.

Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles, 13(4), 997-1027.

Smith, B., Prentice, I.C. & Sykes, M.T. 2001. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Global Ecology & Biogeography 10: 621-637.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences, 11, 2027-2054, 2014.

Zaehle, S., Sitch, S., Smith, B. & Hatterman, F. 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochemical Cycles 19: 3020.

Zaehle, S., Medlyn, B.E., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hickler, T., Luo, Y., Wang, Y.-P., El-Masri, B., Thornton, P., Jain, A., Wang, S., Warlind, D., Weng, E., Parton, W., Iversen, C.M., Gallet-Budynek, A., McCarthy, H., Finzi, A., Hanson, P.J., Prentice, I.C., Oren, R. & Norby, R.J. 2014. Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies. New Phytologist 202: 803–822.

	AI 280 ppm	AI 450 ppm	Max	Min
Standard	-0.67	-0.96	4.7	-11.5
Absent-Absent = 1 (default = 0)	4.43	4.06	10.5	-11.5
Dominant-Dominant = 1 (default =2)	-0.91	-1.13	4.2	-11.5
Both of the above	4.19	3.9	10	-11.5
Minor disagreement = -1, disagreement = -2, major disagreement = -3 (default = 0,-1,-2)	-4.9	-5.23	4.7	-21.5

Table 1. Overall Agreement Index (AI) scores for the 280 ppm and 450 ppm Tortonian runs, as well as the minimum and maximum values calculated with different scores assigned for levels of agreement.

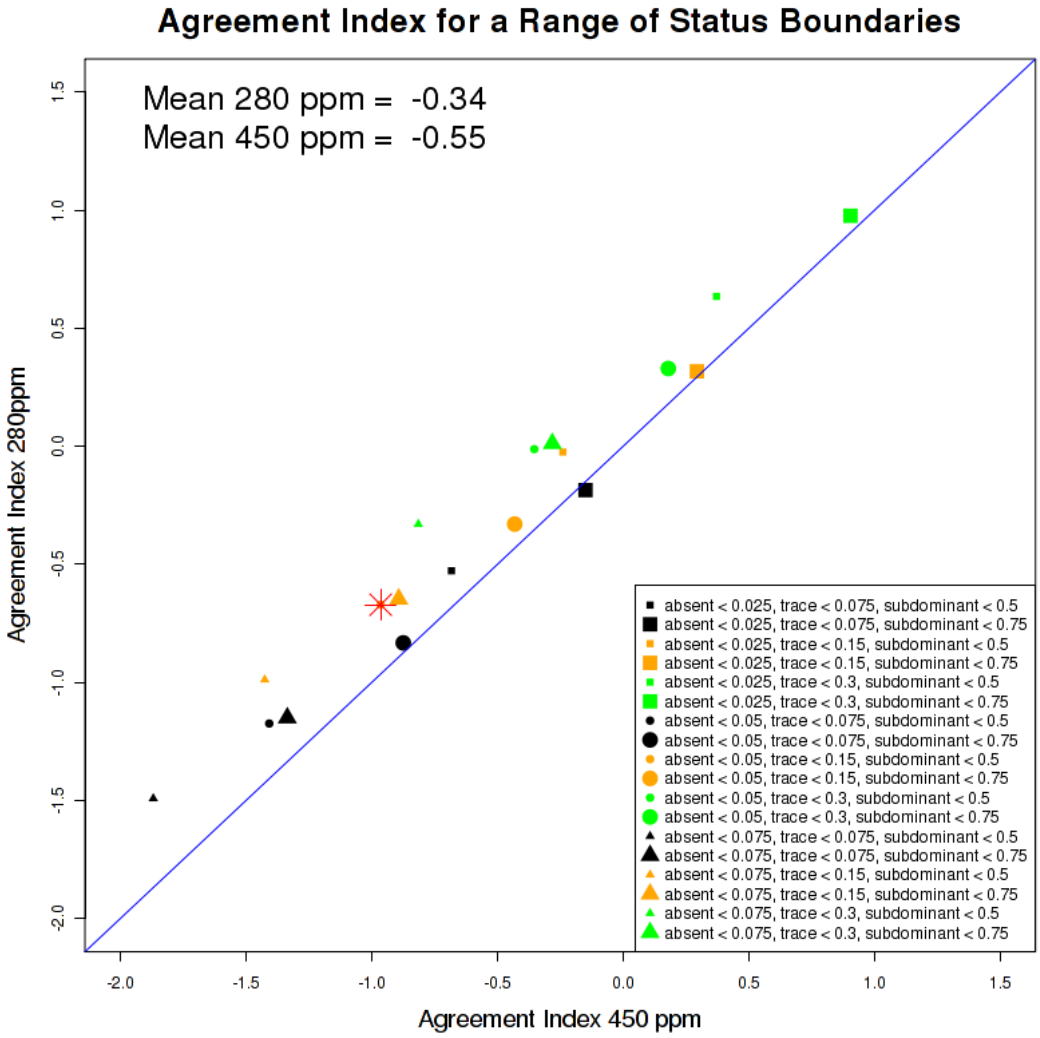


Figure 1. Agreement Index (AI) values for the 280 ppm and 450 ppm runs for different fractional boundaries of the AI statuses.

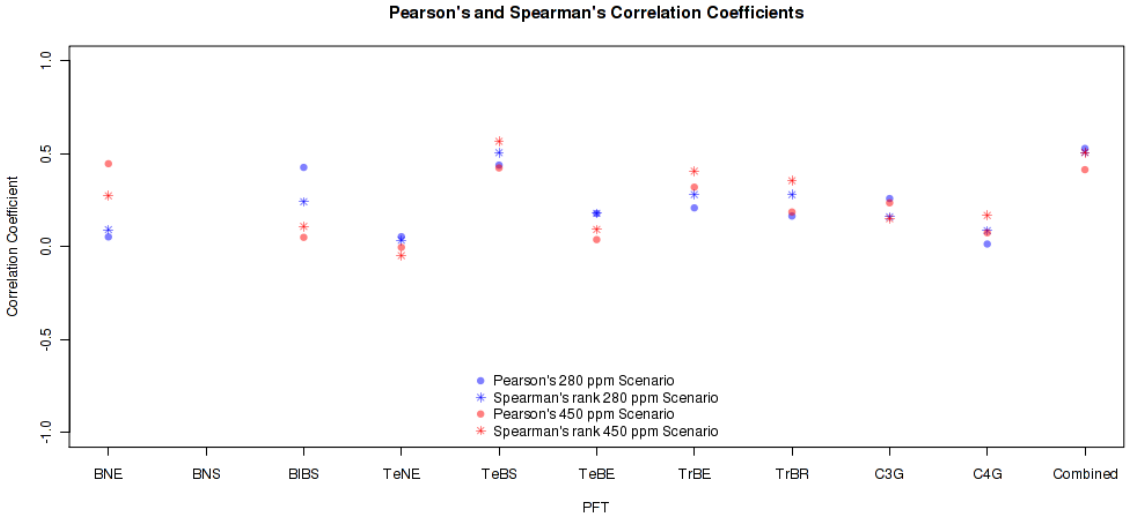


Figure 2. Pearson's product moment correlation coefficient and Spearman's rank correlation coefficients between the paleobotanical data diversity fractions and the simulated LAI fractions for the 280 ppm and 450 ppm CO_2 Tortonian scenarios.

List of changes

Main manuscript

- In section 3.3 a paragraph has been added discussing existing evaluation of LPJ-GUESS (or LPJ-DGVM)
- In section 3.3 the reasons for changes to LPJ-GUESS compared to the standard version are discussed.
- In section 3.3 text has been added to point the reader to section S3 of the supplementary information for model evaluation and discussion of effect sizes.
- In section 3.4 the reader is referred to section S1 of the supplementary material for discussions of Spearman's rank and Pearson's product moment correlation coefficients.
- A new section, section 3.4.1, has been added to discuss in detail previous approaches for comparing fossil data and model output
- A new section, section 3.4.2, has been formed to describe the Agreement Index method. This is primarily formed from existing text with small addition to provide clearer explanation of the method.
- A new section, section 3.4.2, has been added to discuss estimation of chance agreement found using the AI. The section finishes with a paragraph pointing the reader to section S2 of the supplementary material for robustness checks of the AI method.
- Section 3.5 now contains text discussing the aggregation of schlerophylous and perhumid temperate broad-leaved evergreen trees into one PFT.
- Section 4.1 now mentions Figure 5 (which has been moved from the supplement to the main text).
- Section 4.2 now discusses the *p*-value interpretation of the *Z*-scores.
- Section 4.2 now includes a short discussion of the likely magnitude of CO_2 fertilistion effects in the vegetation model,
- One paragraph has been moved from section 4.2 to 4.3 (a regional discussion of model results) as it is more appropriate in section 4.3
- Section 4.3 now contains a discussion of the expanded Table 2 (regional AI scores) and the text discussed in the previous point.
- Sections 4.3.1-4.3.6 have been altered to remove quantitative discussions of the AI scores where there is insufficient data to merit it.
- Section 5 has been modified to emphasise model uncertainty and moderate the conclusions appropriately.

Main tables

- Table 2 and 3 have been combined and expanded to include AI scores from all continents. Note also that the Central Europe region has been expanded to include more fossil sites compared to the original submission

Main figures

- A new figure (Figure 5) has been added to the main text (previously it was in the supplement) which displays the difference in AI scores between to 280 ppm and 450 ppm simulations spatially.

Appendices

- Appendix B has been expanded to explain the biome classification in more detail and now includes a tables which serves as a complete reference for the classification

Supplementary material

- The supplementary material has been expanded significantly. It now includes a discussion of Pearson's product moment correlation coefficient and Spearman's rank correlation coefficient results, extensive robustness checks of the Agreement Index (AI) method and a discussion of estimating agreement by chance, and a discussion of model uncertainty (based on the present day control vegetation and a potential natural vegetation biome map) and signal size using Cohen's Kappa statistic.
- Former Figure S2 (map showing the differences in AI scores at fossil data sites) of the original supplement have been moved to Fig. 5 of the main text.
- Table S5 has been added to show the mapping from fossil taxa to Plant Functional Types.

Title

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

Authors:

Forrest, M. [†1], Eronen, J.T.* [†1,2], Utescher, T.[1,3], Knorr, G.[4], Stepanek, C.[4], Lohmann, G.[4], Hickler, T.[1,5]

Addresses

[1]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt am Main, Germany

[2]Department of Geosciences and Geography, University of Helsinki, PO Box 64, 00014 Helsinki, Finland

[3]Steinmann Institute, University of Bonn, Nussallee 8, D-53115 Bonn, Germany

[4]Alfred Wegener Institute, Bussestrasse 24, D-27570 Bremerhaven, Germany

[5]Department of Physical Geography, Geosciences, Goethe University, Altenhöferallee 1, D-60438, Frankfurt am Main, Germany

[†]= Equal author contribution

* = Corresponding author

Abstract

There is increasing need to understand the pre-Quaternary warm climates, how climate-vegetation interactions functioned in the past, and how we can use this information for understanding the present. Here we report vegetation modelling results for the Late Miocene (11-7 Ma) to study the mechanisms of vegetation dynamics and the role of different forcing factors that influence the spatial patterns of vegetation coverage. One of the key uncertainties is the atmospheric concentration of CO_2 during past climates. Estimates for the last 20 million years range from 280 ppm to 500 ppm. We simulated Late Miocene vegetation using two plausible CO_2 concentrations, 280 ppm CO_2 and 450 ppm CO_2, with a dynamic global vegetation model (LPJ-GUESS) driven by climate input from a coupled AOGCM (Atmosphere-Ocean General Circulation Model). The simulated vegetation was compared to existing plant fossil data for the whole Northern Hemisphere. For the comparison we developed a novel approach that uses information of the relative dominance of different Plant Functional Types (PFTs) in the palaeobotanical data to provide a quantitative estimate of the agreement between the simulated and reconstructed vegetation. Based on this quantitative assessment we find that pre-industrial CO_2 levels are largely consistent with the presence of seasonal temperate forests in Europe (suggested by fossil data) and open vegetation in North America (suggested by multiple lines of evidence). This suggests that during the Late Miocene the CO_2 levels have been relatively low, or that other factors that are not included in the models maintained the seasonal temperate forests and open vegetation.

924

925 1. Introduction

926

927 The Late Miocene (11 to 7 Ma) belongs to the late phase of the Cenozoic climate
928 cooling, during which the seasonality of climate in Europe intensified (e.g.
929 Mosbrugger et al., 2005) and landscapes in North America opened (Eronen et al.,
930 2012). In many regions, it was still characterised by warm and humid climatic
931 conditions compared to today (Micheels et al., 2011, Utescher et al., 2011, Eronen et
932 al., 2012, Fortelius et al., 2014). The global continental configuration in the Miocene
933 was generally comparable to the modern situation with some small differences (e.g.,
934 Herold et al., 2008, Micheels et al., 2011). Marine evidence indicates that tropical sea
935 surface temperatures were similar or even warmer than present in the Early to Middle
936 Miocene (e.g., Stewart et al., 2004), and terrestrial equatorial regions were as warm as
937 today in the Late Miocene (Williams et al., 2005; Steppuhn et al., 2006). The polar
938 and Northern regions were warmer during the whole Miocene (e.g., Wolfe, 1994a,b,
939 Utescher et al., 2011, Popova et al., 2012). Similarly, the North Pacific in the Late
940 Miocene was warmer than today (Lyle et al., 2008). CO_2 levels during the Late
941 Miocene can still not be reconstructed with certainty (see e.g. discussion in Beerling
942 and Royer 2011): estimates for the atmospheric CO_2 levels range from 280 ppm to as
943 high as 500 ppm. Recent studies suggest about 350–500 ppm for the Middle Miocene
944 (Kürschner et al., 2008, Foster et al., 2012, Zhang et al., 2013), and around 280-350
945 ppm for the Late Miocene (Zhang et al., 2013, their figure 5). In addition, terrestrial
946 proxy data suggest that during the Late Miocene there was a marked increase in both
947 temperature and precipitation seasonality (Janis et al., 2002, Mosbrugger et al., 2005,
948 Eronen et al., 2010, 2012). Plant-based data evidence that the increase in temperature

seasonality was mainly effective in the middle to higher latitudes (Utescher et al., 2011), while the evolution of precipitation seasonality was strongly region-dependant and variable throughout the late Miocene (Syabryaj et al., 2007; Utescher et al., 2015). Knorr et al. (2011) modelled the impact of vegetation and tectonic conditions on the Late Miocene climate, and showed that the vegetation has a considerable effect on the climate, and that Late Miocene warmth can be modelled with relatively low CO_2 concentrations at pre-industrial level (278 ppmv). Further, LaRiviere et al. (2012) showed that the oceanic state in the Late Miocene was similar to that of Early Pliocene, with a deeper thermocline, high SSTs, and low SST gradients. They further suggested that, based on their data, during the Late Miocene and earlier times CO_2 and oceanic warmth were decoupled because of deeper thermoclines. The tight link between ocean temperature and CO_2 formed only during the Pliocene when the thermocline shoals and surface water became more sensitive to CO_2. Bolton & Stoll (2013) on the other hand suggested that, based on coccolith data analysis, the atmospheric CO_2 concentration decreased during the latest Miocene (7-5 Ma). They also suggested that atmospheric CO_2 content might have been higher (400-500 ppm, based on Zhang et al., 2013) during the Middle and Late Miocene, and that the substantial ocean surface cooling during the last 15 Ma may reflect the global decrease in the CO_2 concentration.

The Late Miocene is a sub-epoch of the Miocene, which is generally dated roughly between 11 to 5 million years. It includes the Tortonian and Messinian stages. The climate and vegetation models we use in this study use the boundary conditions specific for the Tortonian. The Tortonian comprises the time-interval between 11.6 and 7.2 Ma (Gradstein et al., 2004). It corresponds roughly to European mammal

974 units MN9 to MN12, and Vallesian and lower Turolian mammal zones (Steininger
975 1999). The boundary conditions used for the climate model, as well as the proxy data
976 we use, are dated within these time slices. From here on, we just use the term
977 Tortonian to indicate this time period, and refer to the Late Miocene when we discuss
978 trends in more general terms.
979
980 Here we run the dynamic global vegetation model (DGVM) LPJ GUESS (Smith et
981 al., 2001, Sitch et al., 2003, Ahlström et al., 2012) for the Tortonian with two different
982 CO_2 concentrations to investigate the vegetation dynamics during this period. We use
983 climate data simulated for the Tortonian by Knorr et al. (2011) and Knorr and
984 Lohmann (2014), using a fully coupled AOGCM without any flux corrections. We
985 concentrate on whether the DGVM can create and maintain the mid-latitude seasonal
986 vegetation cover in a generally warmer world, as suggested by the proxy data, and on
987 the sensitivity of the vegetation to CO_2 concentration. We compare our results with
988 existing terrestrial proxy data and previous modelling results, and discuss the
989 implications from our results. Our hypothesis is that in order to maintain the seasonal
990 and open vegetation of the Late Miocene, we need low atmospheric CO_2
991 concentration.
992
993 2. Previous model studies
994
995 Several vegetation model runs have been performed previously for the Late Miocene
996 period. One of the first was a BIOME4 model (Kaplan, 2001) run for the Tortonian by
997 Micheels (2003) to interpolate between the vegetation reconstructed by qualitative
998 interpretation of proxy data from palaeobotanical literature. In this reconstruction the

tropical forests expand in the Tortonian, and their margins shift further poleward. Much of Africa was generally characterised by tropical forest vegetation. Accordingly, the Sahara desert was smaller than today and consisted of steppe and open grassland, rather than sand desert. Woodier Tortonian vegetation replaced the present-day's warm-arid desert, semi-desert and grassland regions.

Francois et al. (2006) used the CARAIB model together with the ECHAM4/ML AOGCM to reconstruct the distribution of vegetation and carbon stocks during the Tortonian (7-11 Ma) with different CO_2 levels. The main difference to our model setup is that ECHAM4 was not coupled to a dynamic ocean model, but a mixed layer ocean model. Their Tortonian run with 280 ppm CO_2 showed a general trend of reduction of desert areas worldwide and appearance of tropical seasonal forests in the warm temperate zone of the Northern Hemisphere, between 30° and 50° (figure 4 of Francois et al., 2006). With their 560 ppm CO_2, most deserts disappeared from the continental surface, except for the Sahara. The extent of tropical seasonal forests also appeared to be extremely sensitive to the atmospheric CO_2 level. Francois et al. (2011) further used the CARAIB model to study the Tortonian vegetation in Europe in detail. On average, their standard 280 ppm run is too cool, with too few temperate humid evergreen trees in Southern Europe compared to their proxy data. Also other models (see below) have struggled to reproduce the seasonal forests in Europe that are known to have existed for the last 10 million years (e.g. Agusti et al., 2003, Mosbrugger et al., 2005).

Pound et al. (2011) used BIOME4, driven by the HadAM3 atmosphere-only general circulation model, and palaeobotanical proxies to create an advanced global data–

model hybrid biome reconstruction for the Tortonian. In their runs boreal forests reach 80°N, and temperate forests were present north of 60°N. Warm–temperate forests cover most of Europe, North America and South-East Asia. There is temperate savannah in central USA. Most areas that are deserts today are covered by grasslands and woodlands in their run. The extent of tropical forests in South America was reduced. Scheiter et al. (2012) used the adaptive DGVM (aDGVM) forced with climate data from HadCM3L and carried out factorial vegetation model runs to investigate the role of fire, emergence of C_4 photosynthesis, and atmospheric CO_2 levels in the vegetation dynamics of Africa. In their runs vegetation openness is mainly determined by fire, generally too much forest cover is simulated if fire disturbance is switched off. The biome pattern is relatively insensitive to changes in the CO_2 concentration or the introduction of herbaceous vegetation with C_4 photosynthesis.

3. Methods

3.1 Palaeoclimate Simulations

The climate simulations have been performed with an AOGCM. The atmosphere model component ECHAM5 (Roeckner et al., 2003) was used at T31 resolution (~3.75°) with 19 vertical levels. The ocean model MPIOM (Marsland et al., 2003) was run with a bipolar curvilinear GR30 resolution (~3°x1.8°) with 40 vertical layers. This modelling approach has been evaluated with proxy data in investigations of the Tortonian (Micheels et al., 2011, Knorr et al., 2011) and the Middle Miocene climate transition (Knorr and Lohmann, 2014). We used the same boundary conditions as

Micheels et al. (2011) with respect to the tectonic setting and the vegetation distribution. We applied minor land-sea modifications, as described in Knorr et al. (2011), e.g., a closed Hudson Bay (Smith et al., 1994). We used data from two model runs with different CO_2 settings, one with a lower CO_2 concentration of 278 ppm (after this referred to as "280 ppm run", from Knorr et al., 2011) and one with a higher CO_2 concentration of 450 ppm (after this referred to as "450 ppm run", from Knorr and Lohmann, 2014).

For further details of the AOGCM model configuration and the boundary conditions we refer the reader to Micheels et al. (2007, 2011), Knorr et al. (2011), and Knorr and Lohmann (2014).

3.2 Correction of present-day biases in climate simulations

To correct for biases in climate simulations, the difference between the Tortonian climate simulations and the pre-industrial control simulation in Knorr et al. (2011) (the Control) was applied to present day climate data to form the palaeoclimate. The Princeton Global Forcing dataset (PGF, Sheffield et al., 2006) was selected as the present day climate baseline. This dataset is a reanalysis product (produced by running an atmospheric circulation model with data assimilation using meteorological measurements) and has been bias-corrected using ground and satellite observations of meteorological variables. Thus it provides global data on a daily or sub-daily time-step which has been dynamically interpolated from station measurements and, by using observed meteorological measurements, is corrected for biases originating from the atmospheric circulation model.

The palaeoclimate anomalies were calculated using the mean values from 100 years of climate simulation and applied following the approach of François et al. (1998) but on a daily, rather than a monthly, time step. The years 1951-1980 were selected to represent the pre-industrial climate, as they give a reasonable compromise between the need for low atmospheric CO_2 (to better represent pre-industrial climate) and the need for maximal instrumentation to measure the climate and so better constrain the atmospheric circulation model.

3.3 Vegetation Simulations

The palaeoclimate model results were used to drive the DGVM LPJ-GUESS. The soil texture map used in the vegetation simulations was derived by translating the soil texture map used by the palaeoclimate AOGCM simulations to the soil classes detailed in Sitch et al. (2003). The representation of vegetation in the palaeoclimate AOGCM comprised statically prescribed land surface classes from Micheels (2003) and as such cannot vary to reach equilibrium with the climate. By using a DGVM with offline climate data we allow the vegetation to reach equilibrium with the (now static) climate. This forms the first step of an asymmetric, iterative offline coupling. Thus we consider our vegetation map to be an iteratively improved version of the original land-cover map of Micheels (2003), improved in the sense that it has undergone one cycle of simulated climate-land surface feedbacks, and has used a more fully developed DGVM with more detailed process representations.

LPJ-GUESS (Smith et al., 2001) combines the generalized representations of the physiological and biophysical processes embedded in the widely used global model LPJ-DGVM (Sitch et al., 2003) with detailed representations of tree population dynamics, resource competition and canopy structure, as generally used in forest gap models (Bugmann 2001, Hickler et al., 2004). LPJ-GUESS (and the closely related LPJ-DGVM model) has been benchmarked against various observations including, for example, NPP (e.g. Zaehle et al., 2005; Hickler et al., 2006), modelled PNV (Hickler et al. 2006; Smith et al. 2014), stand-scale and continental-scale evapotranspiration (AET) and runoff (Gerten et al., 2004), vegetation greening trends in high northern latitudes (Lucht et al., 2002) and the African Sahel (Hickler et al., 2005), stand-scale leaf area index (LAI) and gross primary productivity (GPP; Arneth et al., 2007), forest stand structure and development (Smith et al., 2001, 2014; Hickler et al., 2004), global net ecosystem exchange (NEE) variability (Ahlström et al. 2012, 2015) and CO2 fertilisation experiments (e.g. Hickler et al. 2008; Zaehle et al. 2014; Medlyn et al. 2015).

Here, we build upon a recent version, including a representation of wildfires (Thonicke et al., 2001), the hydrology scheme from Gerten et al. (2004), and updates, in particular concerning the Plant Functional Type (PFT) parameterization described by Ahlström et al. (2012). The bioclimatic limits from Ahlström et al. (2012) were revisited and modified follow the original values in Sitch et al. (2003)., This was motivated by an artefact found in the parameters of Ahlström et al. (2012) whereby in certain areas it was too warm for temperate trees to establish, but too cold for tropical trees. This resulted in treeless belts in South China, Argentina and Florida (see Smith et al. 2014, Figure 2(C) for the model version which does not include nitrogen

limitation). The updated bioclimatic parameters corrected this, but did not result in any other significant differences. ~~as described below. The new bioclimatic limit parameterizations improve the simulated present-day vegetation compared to an independently derived expert map. In our version, the bioclimatic limits follow the original values in Sitch et al. (2003).~~ The boreal/temperate shade-intolerant summergreen broadleaved tree (IBS) PFT in Ahlström et al. (2012) was split into separate boreal and temperate PFTs with temperature limits on photosynthesis, as the other boreal and temperate PFTs, respectively. ~~-~~A Temperate Needle-leaved Evergreen PFT (TeNE) was added based on a similar PFT in Sitch et al. (2003). Both these changes we made to match the PFTs simulated with those classified from the fossil data. The base respiration rates of boreal PFTs were increased compared to temperate trees (as in Hickler et al., 2012), reflecting the general increase of base respiration rates with decreasing temperature (Lavigne and Ryan 1997). ~~Finally, a Temperate Needle-leaved Evergreen PFT (TeNE) was added based on a similar PFT in Sitch et al. (2003).~~ Note that the C_3 and C_4 grass PFTs include forbs, not only grasses. In this paper we refer to these PFTs as grasses because grasses comprise most of the biomass of these PFTs, and this term is more consistent with the terminology used in the palaeobotanical reconstructions. A full list of PFTs and parameter values is given in Appendix A.

The fire model GlobFIRM (Thonicke et al., 2001) with an updated parameterisation as described in Pachzelt et al. (2015~~*in press*~~), but applied globally, was used to simulate wildfires. Representation of fire processes is important when studying vegetation dynamics and structure, particular when considering landscape openness.

We performed a biomisation on the vegetation model output (based on Hickler et al. (2006) but with small changes, see Appendix B) to visualise the simulated Tortonian vegetation (Figure 1a and c), and to compare the vegetation simulation using the PGF climate forcing data ~~of~~ for the present day to a present-day biome map. ~~(Figure S1).~~ These results are presented in section S3 of the supplementary material, where an examination of the model setup's ability to distinguish between present day and Tortonian vegetation can also be found. ~~The pre-industrial control run (Knorr et al., 2011) reproduced the modern biomes (Figure S1a) reasonably well.~~

3.~~4~~3 Statistics to compare modelled and fossil vegetation

Quantitative comparisons of fossil data and model output are challenging. ~~-~~As described below, the palaeobotanical record provides the presence of fossil taxa at a given site and each taxon is then assigned to a PFT. The final values for each site are therefore the number of taxa assigned to each PFT. This is a measure of PFT *diversity*, but typically it is PFT *abundances* which are used to describe vegetation and biomes on a global scale, and it is these quantities, which are provided by vegetation models. There are various difficulties when attempting to draw conclusions from comparisons between diversity data from the fossil record and modelled abundances or biomes. Firstly, abundances and diversity are not necessarily closely correlated; some PFTs might have few taxa but massive abundance (for example Boreal Needle-leaved Trees). Secondly, the fossil record has biases; some PFTs fossilise at higher rates than others, and time-dependent climate fluctuations (Milankovic cycles and the formation and destruction of microclimates) may make the fossil record unrepresentative of PFT diversities over the whole time period. A

1173 further problem is that it is difficult to know how PFT diversities in the fossil record
1174 correlate to an abundance measure that can be simulated by a vegetation model. An
1175 example of a commonly used abundance measure from vegetation models is Leaf
1176 Area Index (LAI), that is the leaf area per unit ground area. Standard statistical tests,
1177 such as Spearmans's rank correlation and Pearson's production moment correlation
1178 coefficient,~~goodness of fit~~ between modelled PFT LAI fraction and the PFT
1179 diversities in the fossil record, did not yield useful results ~~(data not shown)~~, possibly
1180 for the reasons discussed above. These results are shown and discussed in section S1
1181 supplementary material.
1182
1183 3.4.1 Discussion of previous quantitative approaches
1184
1185 To go beyond simple visual comparisons of model and data, and for hypothesis
1186 testing, we require a quantitative measure of agreement between fossil data and model
1187 output. Different approaches have been developed to compare fossil data to model
1188 results with some quantitative element. The study of Pound et al. (2011) uses Cohen's
1189 kappa to determine biome agreement, comparing both the 27 "native" biomes from
1190 BIOME4 and a 7 "megabiome" classification. This does offers a single statistic which
1191 could be used for hypothesis testing. However, there are inherent shortcomings when
1192 using kappa to compare biome classifications and with biome classifications
1193 themselves.
1194
1195 The inherent disadvantage of comparing kappa scores for biomes is that kappa does
1196 not include any mechanism to account for "degrees of difference" which can be
1197 important when considering more than two categories. For example, there is a much

smaller conceptual difference between a “tropical grassland” and a “tropical savanna” than there is between a “tropical grassland” and a “boreal forest”, but that difference is treated identically when calculating Cohen’s kappa. This can be ameliorated to some extent by aggregating to megabiomes as done by Pound et al. (2011), but is inevitably present to some extent. A weighting can also be attempted, but this introduces subjective decisions.

The second argument against comparing potential natural vegetation (PNV) biome distributions using kappa is that PNV biome classifications themselves introduce uncertainty. Potential natural vegetation cannot be measured directly (it no longer exists due to human influence) and so must be reconstructed. There is uncertainty in such reconstructions as evidenced by the differences between PNV biome maps: for example, the horn of Africa is predominantly covered by “tropical deciduous forest” in Haxeltine and Prentice (1996), but is dominated by “dense shrublands” in Ramankutty and Foley (1999). Similarly, the extent of the “tropical deciduous forest” biome in Southern Africa varies considerably between the two maps. Even the biomes categories themselves vary between the maps as different authors make different distinctions. Our experience is that kappa statistics applied to compare different PNV maps can indicate as bad agreement as the one between a model and a PNV reconstruction, when biomes are not aggregated to coarser classes. There are also subjective choices when classifying model output which introduces uncertainty. For example, how much tree LAI or tree cover constitutes a forest? How much for a savanna? The choices for these numbers are not well-motivated and can change the biome boundaries considerably. Concerning the paleobotanical data, we deliberately

did not derive biomes because classifying fossil sites into biomes introduces large uncertainty arising from interpreting the fossil record in terms of vegetation cover.

So whilst comparisons of biomes are clearly useful visual aids and can be a useful cross-check, we decided to use only information on PFT fractions for our main analysis and therefore minimize subjective choices and classifications.~~The approach taken in Salzmann et al. (2008) and Pound et al. (2011) involves classifying both the fossil data and the model output into biomes, which necessarily require subjective choices.~~

The work of François et al. (2011) offers a method for determining agreement between paleobotanical data and simulated vegetation which percentage agreement per PFT based on presence/absence. These per-PFT scores could conceivably be combined to produce overall agreement scores, taking care that PFTs which are mostly absent from the fossil record do not unduly affect the final result. However, the scope of this study is different in nature to that of François et al. The study of François et al. was a regional study with a relatively high degree of taxonomic precision (ie. a more detailed PFT set), whereas this study is global with appropriately coarser taxonomic resolution (ie. a relatively simpler but global PFT set). By means of example, there are 8 purely temperate PFTs in the CARAIB version used in François et al. 2011 compared to only 2 in the default LPJ-GUESS configuration and 4 in the configuration used in our study. Thus by exploiting a high degree of taxonomic precision, presence/absence data were used effectively in the regional study of François et al. In the global study presented here, each PFT spans a much larger geographical extent and there are fewer PFTs at each site for which to make

presence/absence comparison. Thus one would expect the effective differentiating power of such presence/absence to be lesser. So rather than using detailed taxonomic resolution and presence/absence information, we seek to exploit the abundance/diversity fractions which we believe has useful information.

To summarise, for this study, we sought a comparison method which uses abundance/diversity information beyond presence/absence, avoids biomes classifications, avoids Cohen's kappa for multiple categories, and provides a simple number to summarise overall agreement for a given model run.

~~We prefer a metric that uses only the raw data without a biome classification, using more information than provided by presence-absence data, and providing a simple number to summarise overall agreement for a given model run.~~

3.4.2 Calculation of Agreement Index

~~To this end we developed an Agreement Index (AI). This index~~ As motivated above, we developed a novel comparison index which we refer to as the Agreement Index (AI). This index compares the fractional diversity of each PFT at each fossil site (diversity of each PFT divided by the total diversity) to the LAI fraction of that PFT in the corresponding gridcell (LAI for the PFT divided by the total LAI for the gridcell). The LAI values are the growing season maximum values and are averaged over a 30 simulation year period. ~~takes into account all the fractional representations of different PFTs in the model (LAI) and fossil data (number of taxa) for each fossil site.~~ Based on these fractions, each ~~A~~ PFT is assigned ~~can have~~ one of 4 statuses ~~in a~~

~~gridcell~~ in both the fossil data and the model output at each fossil site. ~~-~~These statuses are [fossil, model]: 1) Dominant – fraction in the range (0.50, 1.0], 2) Sub-dominant – fraction in the range (0.15, 0.50], 3) Trace – fraction in the range (0.05, 0.15], 4) Absent – [0, 0.05]. These are then compared between fossil and model for each PFT, and a contribution quantifying the degree of agreement is added to the AI for the gridcell as given in Table 1. The AI is then averaged across all fossil sites.

The logic of the AI is as follows. If a PFT is absent in both the data and the model it contributes 0, since correctly not simulating a PFT is not much of a test of model skill. This also has the desirable effect that a PFT, which is only minimally represented in both the fossil record and the model output, does not strongly affect the final AI value. If the PFT status matches between the model and the data, then it contributes +1, except for if it is the dominant PFT, in which case +2 is added. The dominant PFT is weighted more heavily because it defines the biome and represents the most significant component of the vegetation present. If the model and data mismatch by one category (e.g. the PFT is trace in the model but absent in the data, or dominant in the data but only sub-dominant in the model) then there is a contribution of 0. In such a case the model is not exactly right, but it is not too far away. Given the large uncertainties in inferring relative abundance from fossil diversity data, this degree of statistical mismatch is acceptable. If the data and model differ by two categories (say, the PFT is sub-dominant in the model but absent in the data) this represents a mismatch and contributes -1. Finally, if model and data mismatch by three categories (cases where a PFT is absent in the data but dominant in the model, or vice-versa) a contribution of -2 is added to the AI as this indicates large data-model disagreement.

The range of possible values that the AI can take at a given site is determined by the composition of fossil PFTs at the site. Averaging across all sites used in this analysis gives a range of (-11.4, 4.7). However, this range is relatively meaningless as the chances of getting perfect agreement or perfect disagreement are vanishingly small.

3.4.3 Interpreting Agreement Index scores and quantifying agreement by chance

The Agreement Index method calculates a single score for one model run compared to a fossil dataset. Thus AI scores for two (or more) model runs can be compared and the model run with the highest AI score can be said to have the highest level of agreement with the fossil dataset. This in itself says nothing about the level absolute level of agreement between a particular model simulation and the fossil data (only that one agrees better compared to the other), or about how *much* better one model run agrees with the data than another model run. To address these questions, one requires both an estimate of what agreement could be expected by chance, and an estimate how much variability there is around this value. To quantify this, one can calculate the Agreement Index for a large number of 'random simulations' using a Monte Carlo approach (the exact algorithm to produce these 'random simulations' is important and discussed later). The mean value of these AI scores gives an expectation value for agreement by chance which can be used as a reference point for considering absolute agreement. The standard deviation of these values gives a convenient unit to quantify the typical spread of AI values and indicate how much better a particular model run is compared either to chance agreement or to another model run. Given this standard deviation and mean value, conventional Z scores and *p*-values can be calculated and

interpreted, but the interpretation must always consider the method by which agreement by chance was quantified.

There is no obvious and ubiquitous method to produce a 'random simulation' and various possibilities could be conceived. A truly random simulation would result in unrealistic PFTs combinations and would not be an informative baseline. We chose to construct a 'random simulation' by matching a randomly selected modelled gridcell (from either the 280 ppm simulation or the 450 ppm simulation) to each fossil data site. Because this approach uses model output, it samples the climate space in a fairly even way and simultaneously ensures ecologically realistic PFT combinations. It is therefore a reasonably 'strict' method compared to a more random method. Other approaches for quantifying agreement by chance are tested and discussed in Section S2 of the accompanying supplementary material. We calculated the AI scores for 25,000 'random simulations' using this method. The mean value of these scores was found to be -1.96 which is close to the centre point of the theoretically possible range. The standard deviation was 0.17.

~~In order to simulate the level of agreement that might be expected simply by chance, a set of 10,000 AI values were produced by matching each fossil sites to a randomly selected gridcell chosen from the 280 ppm and 450 ppm model runs combined. This gives an approximate null model with an expectation value for chance agreement and a standard deviation to test for significance. The expectation value was -1.96 (close to the centre point of the theoretically possible range) with a standard deviation of 0.17.~~

3.4.4 Robustness of Agreement Index.

The robustness of the AI was assessed with respect to the subjective choices of the method. Specifically, the choice of boundary values for AI statuses, score assigned for degree of similarity/dissimilarity and random agreement model were all varied and the results are reported in section S2 of the supplementary material. The method showed only limited sensitivity to these choices and no change was large enough to affect the scientific conclusions. We therefore suggest this approach as a robust and quantitative comparison of similar model setups for hypothesis testing, as well as a general measure of agreement between fossil data and simulation results.

3.5~~4~~ Palaeobotanical data

The plant data we used are taken from the NECLIME data set as published in the PANGAEA database (doi:10.1594/PANGAEA), completed by data from the authors (full list of sites is provided in table S4 in the supplementary material). After removing sites with more than 20% aquatic taxa, representing azonal sites (not by macroclimate but by local topographic features determined vegetation, such as riparian vegetation, which is not represented by the vegetation model), the set comprised a total of 167 macro (fruits and seeds, leaves) and micro (pollen/spores) floras, dated to the Late Miocene (11 - 7 Ma). To assign PFTs to the fossil plant record, we classified the Nearest Living Relatives of the fossil plant taxa in terms of PFT types that are used in LPJ-GUESS (see table S5 in the supplementary material~~A1~~). Depending on ecological amplitude of a taxonomic unit and the achievable taxonomic resolution, respectively, a single fossil taxon may represent various different PFTs. Therefore, a matrix containing modern taxa and PFT scores was first established, with PFT scores for each taxon adding up to 1. Diversities of

PFTs were then calculated for all sites by using a matrix with taxa records together with a matrix containing the scores of the represented PFTs. Taxa diversity in the considered floras is highly variable, ranging from 7 to 129, and the floral data set is heterogeneous regarding its representativeness with respect to PFTs and the spatial scales at which palaeovegetation is mirrored (Utescher et al., 2007). Pollen floras usually allow characterizing regional vegetation, while leaves involve a local signal. Regarding the representativeness of fossil data with respect to PFTs, leaf floras reflect arboreal PFTs well, while remnants of herbaceous PFTs and grasses are rarely preserved. In pollen floras, on the other hand, the herbaceous vegetation tends to be over-represented while fruit and seed floras may be biased regarding the richness of aquatics. With all these uncertainties, we decided to use all palaeofloras for maximal geographic coverage, excluding aquatic ones, dated to the studied time slice.

Various PFTs present in the fossil record, such as forbs, shrubs, lianas, tuft trees, aquatics, etc., are not considered in the analysis because they do not have any corresponding PFTs in the model, and therefore cannot be used for proxy data – model inter-comparisons. In Europe, for example, a shortcoming of the applied model version is that it does not distinguish sclerophyllous drought-adapted and laurophyllous perhumid evergreen temperate trees. A sclerophyllous evergreen PFT had been implemented in a model version including the hydraulic architecture of plants (Hickler et al. 2006), but the more general temperate evergreen PFT used here corresponds more closely with the predominantly non-sclerophyllous vegetation of the late Miocene (see Hickler et al. 2006 for details). Herbaceous PFTs occurring in the fossil record were combined with C_3 grasses. Moreover, deciduousness of sites may be over-estimated in the proxy data set, mainly for two reasons. Firstly, many of

the studied floras and obtained PFT spectra have a relatively strong azonal imprint, because they represent riparian vegetation usually common in a subsiding depositional area. Riparian associations in general have a low diversity of evergreen woody species, compared to the zonal vegetation thriving in the same climate. This effect will be suppressed, but not eliminated, by the removal of sites with more than 20% aquatic taxa, as discussed above. Secondly, high scores for the broadleaf-evergreen component are rarely obtained for mid-latitudinal palaeofloras, if taxonomic resolution is limited, because the majority of temperate genera comprise both deciduous and evergreen species.

4. Results and Discussion

4.1. General patterns

The Late Miocene vegetation patterns are broadly similar to the modern day, with the same general pattern, but northward shifts of biomes (Figure 1a, b). The 450 ppm run is overall warmer and wetter, with largest differences found at the mid-latitudes, where tropical and subtropical components have a wider distribution (Figure 1b). A poleward shift of the C_3/C_4 grass boundary at higher CO_2 is evident from the dominant PFT maps (Figure 1c, d), as C_4 photosynthesis is favoured at low atmospheric CO_2 concentrations and at high temperatures (Ehleringer et al., 1997, Sage 2004).

North America is of particular interest in this analysis due to the opening of landscapesthat is documented in proxy data. Although there is scarce botanical

evidence from North America, other proxy sources, like fossil mammals (Janis et al., 2004, Eronen et al., 2012) and phytoliths (e.g. Strömberg, 2011) point strongly to the opening of landscapes during the Miocene. In the 280 ppm run the vegetation of the Great Plains and Rocky mountain area of North America are more open than in the 450 ppm run, and C3 grasses are the dominant PFT over a much larger area (Figure 1a,b). Another region of interest is Europe, because of its high density of palaeobotanical proxy data. Whilst both runs show Europe to be mostly forested, with the expected northwards shift of biome boundaries compared to the present day, the 280 ppm run shows more deciduous vegetation in Central Europe and more open vegetation in the south which agrees better with European proxy data. Figure 5 shows the difference in AI values at all fossil sites, and the better agreement of the 280 ppm run in central Europe due to a relatively larger abundance of deciduous trees is clearly visibly. These results are discussed further below.

One feature that is very different between our model-based reconstructions, and also between different vegetation and climate models, is the vegetation of Greenland (e.g. Francois et al., 2006, Pound et al., 2011, our results). In most cases, Greenland is assumed to be largely covered with taiga and cold deciduous forests instead of the present-day's ice cover, but there is no fossil data to confirm this. Another large-scale feature of note is that the modern-day Sahara region is vegetated with dry grasslands.

4.2 Comparison of 280 ppm and 450 ppm simulations

Our simulation results with both CO_2 concentrations correspond well with other vegetation modelling and reconstruction results (e.g. Francois et al., 2006, 2011,

Pound et al., 2011) and the palaeobotanical data. Using our quantitative approach, we see that the 280 ppm run shows better agreement with palaeobotanical data than the 450 ppm run. Specifically, the 450 ppm reconstruction yields an AI value of -0.97, ~~and a Z-score of 5.8,~~ whereas the 280 ppm reconstruction shows better agreement with an AI value of -0.67. When using the method of quantifying chance agreement described in Section 3.4.3, the 450 ppm reconstruction gives a *Z*-score of 5.8 (Figure 2). ~~,~~The interpretation of this Z-score is that there is $p < 10^{-8}$ probability of randomly selecting 167 modelled gridcells which agree better with the fossil data better than the 450 ppm scenario. The 280 ppm simulation yields ~~and a~~ Z-score of 7.5 (Figure 2~~)).~~, which is 1.7 standard deviations better than the 450 ppm run, and corresponds to $p < 10^{-13}$ probability of getting better agreement by chance.

In order to disentangle the indirect effect of CO_2 on vegetation via climate, and the direct effect of CO_2 on vegetation, we performed additional simulations with 450 ppm CO_2 in the vegetation model with the 280 ppm CO_2 climate model results and vice versa. The vegetation results with 450 ppm climate and 280 ppm vegetation have the worst agreement, with an AI score of -1.02. The run with 280 ppm climate and 450 ppm vegetation yields an AI of -0.60, which is slightly better than the full 280 ppm run. AI scores with the same CO_2 in the climate simulation but different CO_2 in the vegetation simulation are similar, whereas AI scores with different CO_2 in the climate simulation but ~~identical~~ the same CO_2 in the vegetation simulation are more dissimilar (Table 2). ~~-~~Furthermore, the modelled response of vegetation to higher atmospheric CO_2 without nitrogen limitation most likely overestimates CO_2 fertilisation (see e.g. Hickler et al. 2015). So the CO_2 fertilisation seen in the 450 ppm simulation here can be considered to be at the upper bound of the likely effect of a an atmospheric CO_2

concentration of 450 ppm. These factsis strongly suggests that climate CO_2 is the dominant effect in our simulations. The overall effect of CO_2 concentration in the Tortonian simulation is examined further using Cohen's kappa statistic in section S3 of the supplementary material.

~~We see that with 280 ppm in the climate there are more open conditions in North America, regardless of the vegetation CO_2 (Figures 1, 3 and 4). This is strongly supported by fossil mammal and phytolith data (see below). In Central Europe, the tendency towards more deciduous vegetation is also driven by low CO_2 in the climate, not low CO_2 in the vegetation, shown by the Central European AI values in Table 3. In other areas the patterns are less clear. In tropical regions, the direct effect of CO_2 on vegetation is stronger than the effect via climate, possiblye because in these areas temperature and precipitation is not limiting. In cooler areas (in particular the boreal zone), the effect of CO_2 in the climate system of increasing temperatures is stronger than the CO2 fertilisation effect on vegetation, since these areas are temperature limited.~~

The result that 280 ppm run agrees better with the palaeobotanical data poses a question: how can we have the combinations of moderately low CO_2, seasonal mid-latitude conditions, a generally warmer world, and shallower latitudinal temperature gradient at the same time? Generally, so far the answer has been that the CO_2 concentration must have been higher in the past to create the Late Miocene warmth (see introduction). However, there has been increasing evidence that atmospheric CO_2 during the Late Miocene has not been much higher than during pre-industrial times

(e.g. Pearson and Palmer, 2000, Beerling and Royer, 2011, Zhang et al., 2013). This remains an open question, but it is outside the scope of the present study.

4.3 Regional comparison between model runs and palaeobotanical proxies

Regional AI scores are presented alongside the global AI scores in Table 2 (see also Fig. 5 for the difference in AI scores between the 280 ppm and 450 ppm simulations plotted spatially). In the two regions with most fossil sites, Europe and Asia, we see higher AI scores for the 280 ppm run than for the 450 ppm run. In the other regions there are few data points and no clear difference between the CO_2 scenarios. Examining the spatial patterns on a regional level, ~~W~~we see that with 280 ppm in the climate simulation there are more open conditions in North America, regardless of the ~~vegetation~~ CO_2 concentration in the vegetation simulations (Figures 1, 3 and 4). This is strongly supported by fossil mammal and phytolith data (see below). In Central Europe, the tendency towards more deciduous vegetation is also driven by low CO_2 in the climate, not low CO_2 in the vegetation, shown by the Central European AI values in Table 2~~3~~. ~~-~~In other regions~~areas~~ the patterns are less clear. In tropical regions, the direct effect of CO_2 on vegetation is stronger than the effect via climate, possibly because in these areas temperature and precipitation is not limiting. In cooler areas (in particular the boreal zone), the effect of CO_2 in the climate system of increasing temperatures is stronger than the CO2 fertilisation effect on vegetation, since these areas are temperature limited.

4.3.1. Europe

In Europe, ~~the overall agreement between the palaeobotanical data and vegetation simulated with the 280 ppm scenario is better than with the 450 ppm scenario (Figure 5S2). There appear to be two reasons for this, both related to increased seasonality and openness. Firstly,~~the ~~-~~280 ppm CO_2 model run produces more deciduous and less evergreen vegetation in Central Europe and southeastern Europe. Here, the proxy data indicate a stronger tendency for temperate broadleaved deciduous forest (Central Europe), and mixed mesophytic forests (SW Europe, Paratethys realm and E Medit.) (Utescher et al., 2007) and increased seasonality (see also Mosbrugger et al., 2005). This is reflected in the higher AI scores for the 280 ppm run compared to the 450 ppm run (Table ~~3~~2, Figure 5~~S2~~). ~~Secondly, in the 280 ppm run, B~~both the Iberian Peninsula and modern day Turkey are more open in 280 ppm run, with C_3 grasses dominating, which better matches the palaeobotanical data. ~~Among the Iberian sites studied, ca. 50 % can be interpreted to represent a more open vegetation type, for the eastern Paratethys and E Mediterranean, more than 2/3 of the palaeofloras have PFT spectra indicative for more open conditions, ~~These conclusions are also supported by fossil mammal data (e.g. Fortelius et al., 2014).

~~On a more detailed level, the~~In the 280 ppm run ~~depicts a mix of forests in Europe, with temperate deciduous forest in Central Europe and temperate evergreen forests in South-western Europe (Figure 1). A~~a mix of evergreen forests,~~-~~ grasslands and dry savannas covers most of the Mediterranean and areas up to the Caucasus, with varying degrees of openness (Figure 1 and 3). Central and Northern Europe are covered by temperate seasonal forests and boreal forests (Figure 1 and 4). In the 450

ppm run, the temperate evergreen forests become more dominant in Southern Europe and parts of Central Europe compared to the 280 ppm run. The Mediterranean is still a mix of grasslands, savannas and forests, but with a tendency towards the woodier biome types and an increase in temperature evergreen trees (Fig.~~ure~~ 1).

When comparing to other reconstructions and palaeobotanical data it should be noted that, based on proxy data, the late Miocene vegetation in the lower latitudes of Europe has been characterized as Mixed Mesophytic Forest, an association of thermophilous broadleaved summergreens and conifers as canopy trees, with variably diverse evergreen woods in the understory (Utescher et al., 2007). This characteristic type, however, cannot be resolved in the biome system we presently use.

Compared to our results,~~The~~ Pound et al. (2011) BIOME4 simulation produced~~s~~ tropical xerophytic shrublands for Western and Southern Europe. This is a drier vegetation type than the fossil data, and different from our model run. For Central Europe, the BIOME4 simulation exhibits warm mixed forests, and this agrees well with data and our simulations. The Pound et al (2011) simulations also agree in that the boreal forests are confined to the extreme north of Europe.

The 200/280 ppm global simulations of Francois et al. (2006) produce vegetation in Europe which is very similar to the present day, whereas the 560 ppm run produces tropical seasonal forests in Europe. The presence of tropical seasonal forests in Europe is not well-supported by palaeobotanical proxy data. All of their simulations show a greater extent of the boreal forest than in either in Pound et al. (2011) or our simulations.

In the higher resolution, regional study of Francois et al. (2011), most of Europe is dominated by cool-temperate mixed and temperate broadleaved deciduous forests, but there are ~~is presence of~~ warmer vegetation types present around the Adriatic Sea and in the north of Turkey. Warm-temperate mixed forests grow around the western part of the Paratethys, and an extension of the tropical grassland around the Mediterranean Sea can be observed. These latter aspects are similar to our simulations.

4.3.2 North America

Our 280 ppm model run exhibits vegetation that is similar to the present day in North America. Compared to the 450 ppm runs, this vegetation is more open and seasonal in the Great Plains and Rocky Mountains. The openness is apparent from the increase of C_3 grass PFT dominance, and from the reduction of tree cover and the corresponding savanna classification in the biome plots (Figure 1c,d; Figures 3 and 4). The increased seasonality is shown by the reduction in dominance of the temperate broadleaved evergreen PFT, and by the increase of C_3 grass at the expense of trees. Whilst there are few fossil data points in North America, other available data from isotopes (Passey et al., 2002), mammalian community structure (Janis et al., 2004), mammal-based precipitation estimates (Eronen et al., 2012), as well as phytoliths (Strömberg, 2005) support the open landscapes and graze-dominated faunas during the Tortonian in the Great Plains, as do both midland plant localities in our record (sites Kilgore, Antelope; C3 PFT diversity fraction 20, 60 %). In addition, the data presented in Pound et al. (2011) indicate more open and seasonal vegetation in this region during the Tortonian. In light of these sources of evidence, it appears that the 280 ppm simulation reproduces the vegetation of the central North America

considerably better than the 450 ppm simulation. The importance of low CO_2 for maintaining open landscapes has also been suggested by other modelling studies. Harrison and Prentice (2003), for example, found that the BIOME4 vegetation models consistently overestimated glacial tree cover, if physiological effects of low atmospheric CO_2 were not accounted for. Experimental elevation of CO_2 above ambient levels has been shown to promote shrub encroachment into steppes (Morgan et al., 2007).

A further notable difference is that the 450 ppm simulation exhibits a strong northward movement of biome boundaries compared to the 280 ppm run, which are indicative of a considerably warmer and wetter climate (Figure 1a, b). There is a northward shift of the boreal/temperate boundary in the 450 ppm run compared to the 280 ppm run. Temperate forests have larger extent, and treeline shifts northwards, almost completely replacing tundra in the higher latitudes. In similar fashion, evergreen trees dominate larger areas than deciduous trees in the temperate coastal forests, which may also be linked to the seasonality and humidity changes mentioned above.

In the Southwest and near the Gulf of Mexico, the results are similar in 280 ppm and 450 ppm runs. In the Southwest and south of North America, both simulations produce dry and open vegetation that is similar to the present day (Figure 1a,b). The runs indicate xeric woodlands and shrublands, dominated by temperate evergreen trees. Further north, these biomes transition to temperate deciduous forests along the Eastern Seaboard, which is in broad agreement with the proxy-based results obtained from the Pacific coastal sites between 35 and 45 °N. The main difference between the

280 ppm and 450 ppm runs is that the transitions occur further north in the 450 ppm simulation.

Compared to Pound et al. (2011), in North America our 280 ppm run produces much more open vegetation in the Great Plains, whereas Pound et al. (2011) find more forests. In addition, Pound et al. (2011) reconstruct a large band of temperate grasslands that replaces northern temperate and boreal forests. This is also seen in their Asian reconstruction at similar latitudes, but is not seen in any other reconstruction.

Our model results are fairly consistent with the François et al. (2006) CARAIB model results (their 280 ppm standard Tortonian run). The main differences from our results in North America are that we produce much more open vegetation with 280 ppm CO_2, and much of their eastern forests are tropical seasonal forests, indicating warmer climate. The low CO_2 run of François et al. (with 200 ppm), on the other hand, produced temperate mixed forests in much of North America, with only western North America being more open.

4.3.3 Asia

In Asia, the expected northward biome shifts in the boreal/temperate zone is observed in the 450 ppm simulation relative to the 280 ppm simulation. In a similar fashion to North America and Europe, the temperate-boreal boundary and treelines are at higher latitudes with higher CO_2, resulting in a larger area of temperate deciduous forest, and almost no tundra or boreal deciduous forest, in the 450 ppm simulation (Figure 1a, b).

The 280 ppm biome boundaries are approximately similar to the present day, with the exception that the temperate deciduous forest encroaches much further from Europe into Asia. ~~The only three proxy data points in boreal Asia (Kamchatka, sites Bayokov H1172, Nekkeiveem H3658, Yanran H3690; mixed broadleaved deciduous-conifer forest and mixed shrubland; cf. Popova et al., 2013) indicate that the 280 ppm run fits slightly better (Figure 5S2).~~

Both simulations exhibit a large grass-dominated steppe in Central Asia, but the landscape is not as open as in the present day vegetation. This grass steppe is larger in the 280 ppm run than in the 450 ppm run, and extends slightly further northwards in the western part (Figure 1a, b). The small difference in aridity and openness in the Asian continental interior between the CO_2 concentration scenarios is much less compared to North America. The few inland proxy points in Central Asia (sites Dunhuang, Kuga Xinjiang, S Junggar, Xining Minhe Basin) all have significantly raised proportions of C3 herb component~~, and indicate reasonable agreement~~, with no difference between the different CO_2 simulations~~, though a considerable broadleaved arboreal diversity in the proxy data points to more forested conditions when compared to the model~~. ~~The coastal points at similar latitude on the East China Sea show better agreement with the 280 ppm run (Figure 1a,b).~~ The 280 ppm run shows more temperate broadleaved evergreen trees in southern and eastern China and the surrounding area, than in the 450ppm run.

~~Consequently, better agreement index scores are present in the 280 ppm run.~~

There are few differences between the 280 ppm and 450 ppm simulations in Southwest Asia, South Asia and Southeast Asia; both produce grasslands in the

western areas and savanna in east. The savanna transitions to tropical forests in the southeast. However, the 280 ppm run produces dryer grasslands in the west, and slightly fewer trees in the east. Furthermore, the evergreen tropical forest of the 280 ppm scenario (and in present day simulations) is replaced by tropical seasonal and tropical deciduous forests in the 450 ppm scenario. This is unexpected and observed in the 450 ppm scenario across the humid tropics, and is discussed further below. There are essentially no proxy data available for comparison in these areas. It is known that the present day simulation underestimates tree cover in these areas, so the palaeo model results should be treated with caution.

The Pound et al. (2011) model/proxy hybrid reconstruction shows a similar boreal range in Asia as the 450 ppm run presented here, but with a large band of temperate grasslands separating the boreal and temperate forests. This band is not seen in our reconstructions, but is also simulated for North America in Pound et al. (2011). Elsewhere, the reconstructions are broadly similar, although the Pound et al. (2011) model has more tree cover over much of Central and East Asia (with savanna being present instead of grasslands, and more temperate forests being present on the east coast) and parts of southern and south-eastern Asia (with more tropical trees). All the vegetation reconstructions of François et al. (2006) have a large area of boreal forest in the north, particularly in the northeast, and regardless of CO_2 concentration. They also show greater abundances of trees in the southeast and less openness in the continental interior compared to our runs, although this difference is less pronounced in their lower CO_2 simulations.

4.3.4. Africa

Both of our Tortonian simulations show grasslands in the modern-day Sahara desert (Figure 1a, b). A green Sahara is consistent with generally warmer global climate (e.g. Micheels et al., 2011, Knorr et al., 2011) and this feature is broadly similar to the reconstruction of Pound et al. (2011), which shows only small areas of desert with large areas of tropical xerophytic shrubland. François et al. (2006) did not reconstruct a green Sahara, and shows some areas that are desert at all CO_2 concentrations. The simulation of Scheiter et al. (2012) also showed a large Sahara desert.

Starting from the equator and moving polewards, both of our simulations exhibit a progression from full tree cover in equatorial Africa, changing to savanna biomes, and finally becoming grasslands with near zero tree cover at ±15°N. This pattern is the same as for the present day. The 450 ppm scenario produces more trees, as would be expected from a more humid world with higher CO_2. The higher CO_2 scenario also favours deciduous tropical trees over evergreens, as can be observed in the other humid tropical forests (Figure 1a,b). The reconstructions of Pound et al. (2011), and of François et al. (2006), all show evergreen tree dominating the most equatorial region with a similar gradient of tree cover, but Pound et al. (2011) transitions to shrublands instead of grasslands. The 280 ppm and 560 ppm CO_2 scenarios of François et al. (2006) feature a much greater extent of tropical deciduous forest in Southern Africa.

At the southern and northern extremes of Africa, limited amounts of woody vegetation appear in both our simulations. In the 450 ppm scenario this vegetation

contains some tropical trees, whereas in the 280 ppm scenario this vegetation is purely temperate.

The Scheiter et al. (2012) simulation with C_4 grasses and fire with 280 ppm (Figure 1i in Scheiter et al. 2012) is extremely close to our simulation result with 280 ppm for Africa, but without a green Sahara. In their runs, there is no perfect agreement between proxy data and any one specific simulation scenario. The best agreement is achieved in simulations with fire at 280 ppm CO_2. Their model run with 400 ppm CO_2 and fire changes the pattern slightly, with more woodland in the tropics, and less tropical evergreen forests. This is similar to our 450 ppm CO_2 run where our tropical evergreen forest cover decreases. Unlike the Scheiter et al. (2012) 400 ppm run, in our high CO_2 run the change is from evergreen forest to raingreen forest. In our simulations the forest fraction in the tropics is larger with higher atmospheric CO_2 concentration. This begets more investigation into the tropical vegetation dynamics during the Miocene. The presently available palaeobotanical data is not sufficient for deriving the general broad-scale pattern of raingreen versus evergreen forest.

4.3.5 South America

In South America our Tortonian results show relatively little change compared to the present-day simulation, with the noticeable exception that the savanna biome of modern day Cerrado is much larger in both the high and low CO_2 Tortonian runs (Figure 1a, b). The southern tip of South America is evidently warmer and more humid in the Tortonian runs, as is apparent from the reconstruction of woody temperate biomes that are dominated by broadleaved evergreen trees, as opposed to

the more open and cooler biomes in the present day simulation. The 280 ppm scenario shows a lower fraction of trees that the 450 ppm simulation.~~, and this more open and xeric vegetation agrees slightly better with the two palaeobotanical data points in Patagonia~~. The tendency for raingreen tropical trees to replace evergreens at higher CO_2 concentrations (as in Africa and Southeast Asia) is also observed.

The Pound et al. (2011) results are similar to the Tortonian runs presented here, and the reconstructions have in common a larger savanna area, and a warmer, more forested southern tip of South America compared to the present day simulations (Figure 1a, b, Figure S1). The François et al. (2006) 280 ppm model predicts much more closed environments for the whole continent, with tropical forest extending also to the south where our model produces moist savannas, and the eastern part being dominated by tropical seasonal forests. They produce a similar output for the 560 ppm run, and even their 200 ppm run has much more forests than either of our model runs.

4.3.6. Australia

In both of our Tortonian model runs, much of Australia is covered by tall grasslands (Figure 1a, b). The south is slightly more arid, with some dry grassland in the 450 ppm scenario, and a greater extent of dry grasslands and some xeric shrublands/steppe in the 280 ppm scenario. Along the northeast coast tropical trees are present, resulting in savanna biomes (Figure 1a,b). It should be noted that the present day simulation does not reproduce the large extent of xeric shrublands/steppe in the present day biome map (Figure ~~1a, b~~S4~~a1~~). This may be due to the lack of any shrub PFTs in the parameterisation of LPJ-GUESS. In contrast, the reconstruction of Pound et al.

(2011) with BIOME4 (which explicitly includes shrubland biomes) does include a large area of tropical xerophytic shrubland in their Tortonian simulation, and some in the present day simulation. Their Tortonian simulation also produces a band of savanna along the north east coast, and elements of temperate forest to the south. These forests are not as widespread as in the proxy data, resulting in large corrections in this area. This is mirrored in our results, as the 450 ppm run, with its larger quantity of temperate trees, agrees ~~slightly better~~ with the limited proxy data available in the South (Figure 1a, b).

The François et al. (2006) 280 ppm model produces grasslands over much of Australia with higher CO_2, and semi-desert and desert with lower CO_2. It also shows a band of tropical seasonal forest vegetation along the northeastern coast which extends considerably further inland at higher CO_2 concentrations. On a general level, all the models produce arid biomes over much of Australia, but their exact distributions differ substantially. This may be due to the different representation of xeric vegetation, particularly shrubs, ~~-~~and due to differences in the classification of biomes, particularly shrublands.

5. Summary and Conclusions

Here, we simulated Tortonian vegetation under two plausible atmospheric CO_2 concentrations, using a dynamic global vegetation model forced by AOGCM-based palaeoclimate simulations. We applied a novel approach for comparing modelled vegetation with palaeobotanical data. This approach allowed us to quantitatively test which CO_2 scenario agreed better with the proxy data.

Our results show that the agreement between modelled vegetation and palaeobotanical data is consistently (i.e. overall and in each world region) higher for the 280 ppm model run compared to the 450 ppm run. In other words, the CO_2 level needs to be moderately low in order to maintain the seasonal and open landscapes that are the hallmarks of Late Miocene environments. ~~This strongly suggests that atmospheric CO_2 levels were relatively low during the Late Miocene.~~

The results are most striking for Central Europe and for Central and West America. The 280 ppm run produces deciduous forests in Central Europe and open landscapes in Southern Europe, in agreement with the palaeobotanical evidence, whereas the 450 ppm run produces more evergreen forests. Similar differences in openness in Central and Western North America occur in the simulations. Due to the scarcity of palaeobotanical data in most of North America, higher AI values cannot be observed for the 280 ppm run. However, the open landscapes observed in the 280 ppm run are supported by multiple lines of evidence, including fossil mammal data, isotopes, and phytoliths. Results from factorial runs, assuming different CO_2 concentrations in the climate and the vegetation model, suggest that climatic effect of CO_2 are most important. Physiological CO_2 effects also play a secondary role, in particular in Central and Western North America. ~~In the continental interior of East Asia there is a small difference in aridity and openness between the two CO_2 concentration scenarios. The few proxy data available inland and in coastal areas along the East China Sea also show better agreement with the 280 ppm run.~~ There are still uncertainties in the models, and these results should be tested with different models~~,~~

~~too~~. Next phase of studies should test our results also using marine data and marine ecosystem models to compare between terrestrial and marine realms.

Our results ~~strongly~~ suggest that atmospheric CO_2 levels were relatively low during the Late Miocene, and that ~~We conclude that~~ the Late Miocene fossil vegetation data can be used in conjunction with vegetation/climate modeling ~~can be used~~ to constrain CO_2 concentrations in the atmosphere~~.~~. ~~Further studies shall test this idea using marine data in connection with marine ecosystem models.~~

Acknowledgments

JTE was supported by A.v Humboldt foundation grant and a Marie Curie fellowship (FP7-PEOPLE-2012-IEF, grant number 329645, to JTE and TH). MF and TH acknowledge support through the LOEWE funding program (Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of Hesse's Ministry of Higher Education, Research, and the Arts. TU thanks the German Science Foundation for the funding obtained (MI 926/8-1). This study is a contribution to *NECLIME* (Neogene Climate Evolution of Eurasia). G.K. and C.S. acknowledge funding by the 'Helmholtz Climate Initiative REKLIM' (Regional Climate Change), a joint research project of the Helmholtz Association of German research centres.

References

Agusti, J., Sanz de Siria, A. and Garcés, M.: Explaining the end of the hominoid experiment in Europe. J. Human Evol., 45, 145-153, 2003

Ahlström, A., Schurgers, G., Arneth, A., & Smith, B.: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. Environmental Research Letters, 7, 044008, 2012.

Ahlström, A., Raupach, M.R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J.P., Friedlingstein, P., Jain, A.K., Kato, E., Poulter, B., Sitch, S., Stocker, B.D., Viovy, N., Wang, Y.-P., Wiltshire, A., Zaehle, S. & Zeng, N. 2015. The dominant role of semi-arid ecosystems in the trend and variability of the land CO_2 sink. Science 348: 895-899.

Arneth, A., Miller, P.A., Scholze, M., Hickler, T., Schurgers, G., Smith, B. & Prentice, I.C. 2007. CO_2 inhibition of global terrestrial isoprene emissions: Potential implications for atmospheric chemistry. Geophysical Research Letters 34: L18813.

Beerling D.J. and Royer D.L.: Convergent Cenozoic CO_2 history. Nature Geosci., 4, 418-20, 2011.

Bolton, C.T. and Stoll, H.M.: Late Miocene threshold response of marine algae to carbon dioxide limitation, Nature, 500, 558-562, 2013.

Bugmann, H.: A review of forest gap models. Climatic Change, 51, 259–305, 2001.

Ehleringer, J. R., Cerling, T. E., and Helliker, B. R.: C4 photosynthesis, atmospheric CO_2, and climate. Oecologia, 112, 285-299, 1997

Eronen, J.T., Fortelius, M., Micheels, A., Portmann, F.T., Puolamäki, K. , Janis, C.M.: Neogene Aridification of the Northern Hemisphere. Geology, 40, 823-826, 2012.

Eronen, J.T., Puolamäki, K., Liu, L., Lintulaakso, K., Damuth, J., Janis, C., and Fortelius, M.: Precipitation and large herbivorous mammals , part II: Application to fossil data. Evolutionary Ecology Research, 12, 235-248, 2010

Fortelius, M., Eronen, J.T., Kaya, F., Tang, H., Raia, P., and Puolamäki, K.: Evolution of Neogene Mammals in Eurasia: Environmental Forcing and Biotic Interactions. Annual Review of Earth and Planetary Sciences, 42, 579-604, 2014

Foster, G. L., Lear, C.H. and Rae, J.W.B.: The evolution of pCO_2, ice volume and climate during the middle Miocene. Earth Planet. Sci. Lett., 341-344, 243 – 254, 2012.

François, L. M., Delire, C., Warnant, P. and Munhoven, G.: Modelling the glacial–interglacial changes in the continental biosphere. Global and Planetary Change, *16*, 37-52, 1998.

François L, Utescher T, Favre E, Henrot AJ, Warnant P, Micheels, A., Erdei, B., Suc, J.P, Cheddadi, R. and Mosbrugger, V.: Modelling Late Miocene vegetation in Europe: Results of the CARAIB model and comparison with palaeovegetation data. Palaeogeogr., Palaeoclim., Palaeoecol., 304, 359–378, 2011.

Francois, L., Ghislain, M., Otto, D. and Micheels, A.: Late Miocene vegetation reconstruction with the CARAIB model. Palaeogeogr., Palaeoclim., Palaeoecol., 238, 302–320, 2006.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance – hydrological evaluation of a dynamic global vegetation model. Journal of Hydrology, 286, 249–270, 2004

Gradstein, F.M., Ogg, J.G., Smith, A.G., Agterberg, F.P., Bleeker, W., Cooper, R.A., Davydov, V., Gibbard, P., Hinnov, L.A., House, M.R. (†), Lourens, L., Luterbacher, H-P., McArthur, J., Melchin, M.J., Robb, L.J., Sadler, P.M., Shergold, J., Villeneuve, M., Wardlaw, B.R., Ali, J., Brinkhuis, H., Hilgen, F.J., Hooker, J., Howarth, R.J., Knoll, A.H., Laskar, J., Monechi, S., Powell, J., Plumb, K.A., Raffi, I., Röhl, U., Sanfilippo, A., Schmitz, B., Shackleton, N.J., Shields, G.A., Strauss, H., Van Dam, J., Veizer, J., Van Kolfschoten, Th. and Wilson, D.: Geologic Time Scale 2004. Cambridge University Press, 2004.

Harrison S. and Prentice C.I.: Climate and CO_2 controls on global vegetation distribution at the last glacial maximum: analysis based on paleovegetation data, biome modelling and paleoclimate simulations. Global Change Biology, 9, 983-1004, 2003.

Haxeltine, A. and Prentice, I. C.: BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. Global Biogeochemical Cycles, 10, 693-709, 1996.

Hickler, T., Smith, B., Sykes, M. T., Davis, M. B., Sugita, S. and Walker, K.: Using a generalized vegetation model to simulate vegetation dynamics in northeastern USA. Ecology, 85, 519-530, 2004.

Hickler, T., Eklundh, L., Seaquist, J., Smith, B., Ardö, J., Olsson, L., Sykes, M.T. & Sjöström, M. 2005. Precipitation controls Sahel greening trend. Geophysical Research Letters 32: L21415.

Hickler, T., Prentice, I. C., Smith, B., Sykes, M. T. and Zaehle, S.: Implementing plant hydraulic architecture within the LPJ Dynamic Global Vegetation Model. Global Ecology and Biogeography, 15, 567-577, 2006.

Hickler, T., Vohland, K., Feehan, J., Miller, P. A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T.R., Cramer, W., Kühn, I., and Sykes, M. T.: Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species- based dynamic vegetation model. Global Ecology and Biogeography, 21, 50-63, 2012.

Hickler, T., Rammig, A. & Werner, C. 2015. Modelling CO_2 impacts on forest productivity. Current Forestry Reports 1: 69-80.

Herold, N., Seton, M., Müller, R.D., You, Y. and Huber, M.: Middle Miocene tectonic boundary conditions for use in climate models. Geochemistry, Geophysics, Geosystems, 9, Q10009, 2008

Janis, C.M., Damuth, J. and Theodor, J.M.: The origins and evolution of the North American grassland biome: the story from the hoofed mammals. Palaeogeogr. Palaeoclimatol. Palaeoecol., 177, 183-198, 2002.

Janis, C.M., Damuth, J. and Theodor, J.M.: The species richness of Miocene browsers, and implications for habitat type and primary productivity in the North American grassland biome. Palaeogeogr. Palaeoclimatol. Palaeoecol., 207, 371-398, 2004.

Kaplan, J.O.: Geophysical applications of vegetation modeling. No. ARVE-THESIS-2009-001. Lund University, 2001.

Knorr G., Butzin M., Micheels A. and Lohmann G.: A warm Miocene climate at low atmospheric CO_2 levels. Geophys. Res. Lett., 38, L20701, 2011.

Knorr, G. and Lohmann, G.: Climate warming during Antarctic ice sheet expansion at the Middle Miocene transition. Nature Geosci., 7, 376–381, 2014.

Kürshner, W.M., Kvacek, Z. and Dilcher, D.L.: The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. PNAS 105, 449-453, 2008.

LaRiviere, J.P., Ravelo, A.C., Crimmins, A., Dekens, P.S., Ford, H.L., Lyle, M. and Wara, M.W.: Late Miocene decoupling of oceanic warmth and atmospheric carbon dioxide forcing. Nature, 486, 97–100, 2012.

Lavigne, M. B. and Ryan, M. G.: Growth and maintenance respiration rates of aspen, black spruce and jack pine stems at northern and southern BOREAS sites. Tree Physiology, 17, 543-551, 1997.

Lucht, Wolfgang, et al. "Climatic control of the high-latitude vegetation greening trend and Pinatubo effect." Science 296.5573 (2002): 1687-1689.

Lyle, M., Barron, J., Bralower, T.J., Huber, M., Olivares Lyle, A., Ravelo, A.C., Rea, D.K. and Wilson, P.A.: Pacifc Ocean and Cenozoic evolution of climate. Reviews of Geophysics, 46, RG2002/2008, 2008.

Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M. and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. Ocean Modelling, 5, 91-127, 2003.

Medlyn, B.E., Zaehle, S., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hanson, P.J., Hickler, T., Jain, A.K., Luo, Y., Parton, W., Prentice, I.C., Thornton, P.E., Wang, S., Wang, Y.-P., Weng, E., Iversen, C.M., McCarthy, H.R., Warren, J.M., Oren, R. & Norby, R.J. 2015. Using ecosystem experiments to improve vegetation models. Nature Climate Change 5: 528-534.

Micheels A, Bruch AA, Eronen J, Fortelius M, Harzhauser M, Utescher, T. and Mosbrugger, V.: Analysis of heat transport mechanisms from a Late Miocene model

experiment with a fully-coupled atmosphere-ocean general circulation model. Palaeogeogr. Palaeoclimatol. Palaeoecol. 304, 337-50 , 2011.

Micheels, A.: Late Miocene climate modelling with echam4/ml – the effects of the palaeovegetation on the Tortonian climate. Unpublished Thesis, Eberhard-Karls Universität Tübingen, 2003.

Micheels, A., A. A. Bruch, D. Uhl, T. Utescher, and V. Mosbrugger: A late Miocene climate model simulation with ECHAM4/ML and its quantitative validation with terrestrial proxy data, Palaeogeogr. Palaeoclimatol. Palaeoecol., 253, 251-270, 2007.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. Ecological modelling, 62(4), 275-293.

Morgan J.A., Milchunas D.G., LeCain D.R., West M. and Mosier A.R.: Carbon dioxide enrichment alters plant community structure and accelerates shrub growth in the shortgrass steppe. PNAS, 37, 14724-14729, 2007.

Mosbrugger, V., Utescher, T. and Dilcher, D.L.: Cenozoic continental climatic evolution of Central Europe. PNAS, 102, 14964–14969, 2005.

~~Pachzelt, A., Forrest, M., Rammig, A., Higgins, S. and Hickler, T.: Potential impact of large ungulate grazers on African vegetation, carbon storage and fire regimes. Global Ecology and Biology, in press~~
Pachzelt, A., Forrest, M., Rammig, A., Higgins, S. I. and Hickler, T. (2015), Potential impact of large ungulate grazers on African vegetation, carbon storage and fire regimes. Global Ecology and Biogeography, 24: 991–1002. doi: 10.1111/geb.12313

Passey, B.H., Cerling, T.E., Perkins, M.E., Voorhies, M.R., Harris, J.M. and Tucker, S.T.: Environmental change in the Great Plains: An isotopic record from fossil horses. J. Geol., 110, 123-140, 2002

Pearson, P. N., and Palmer, M.R.: Atmospheric carbon dioxide concentrations over the past 60 million years, Nature, 406, 695-699, 2000.

Popova, S., Utescher, T., Gromyko, D., Bruch, A. and Mosbrugger, V.: Palaeoclimate Evolution in Siberia and the Russian Far East from the Oligocene to Pliocene – Evidence from Fruit and Seed Floras. - Turkish Journal of Earth Sciences, 21, 315-334, 2012.

Popova, S., Utescher, T., Gromyko, D.V., Mosbrugger, V., Herzog, E., and Francois, L.: Vegetation change in Siberia and the Northeast of Russia during the Cenozoic Cooling – a study based on diversity of plant functional types. Palaios, 28, 418-432, 2013.

Pound, M.J., Haywood, A.M., Salzmann, U., Riding, J.B., Lunt, D.J., Hunter, S. A: Tortonian (Late Miocene, 11.61–7.25 Ma) global vegetation reconstruction, Palaeogeogr., Palaeoclim., Palaeoecol., 300, 29-45, 2011.

Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles, 13(4), 997-1027.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U. and Tompkins, A.: The atmospheric general circulation model ECHAM5—Part I: Model description, Tech. Rep. 349, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 2003.

Sage, R. F.: The evolution of C4 photosynthesis. New phytologist, 161, 341-370, 2004.

Scheiter, S., Higgins, S.I., Osborne, C.P., Bradshaw, C., Lunt, D., Ripley, B.S., Taylor, L.L. and Beerling, D.J.: Fire and fire-adapted vegetation promote C4 expansion in the late Miocene. New Phytologist ,195, 653-666, 2012.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. Journal of Climate, 19, 3088-3111, 2006.

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K. and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology, 9, 161–185, 2003.

Smith, A. G., Smith, D. G., and Funnell, B. M.: Atlas of Cenozoic and Mesozoic coastlines. Cambridge University Press, 1994.

Smith, B., Prentice, I.C. and Sykes, M.T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Global Ecology and Biogeography, 10, 621–637, 2001.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences, 11, 2027-2054, 2014.

Steppuhn, A., Micheels, A., Geiger, G., and Mosbrugger, V.: Reconstructing the Late Miocene climate and oceanic heat flux using the AGCM ECHAM4 coupled to a mixed-layer ocean model with adjusted flux correction. Palaeogeography, Palaeoclimatology, Palaeoecology, 238, 399–423, 2006.

Steininger, F.F.: Chronostratigraphy, geochronology and biochronology of the Miocene European Land Mammal Mega-Zones (ELMMZ) and the Miocene mammal-zones, in: The Miocene Land Mammals of Europe, Verlag Dr. Friedrich Pfeil, Munich, Germany, 9-24, 1999.

Stewart, D.R.M., Pearson, P.N., Ditchfield, P.W. and Singano, J.M.: Miocene tropical Indian Ocean temperatures: evidence from three exceptionally preserved foraminiferal assemblages from Tanzania. Journal of African Earth Sciences, 40, 173–190, 2004.

Strömberg, C.A.E.: Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. PNAS, 102, 11980-11984, 2005.

Strömberg C.A.E.: Evolution of grasses and grassland ecosystems. Annual Reviews of Earth and Planetary Sciences. 39, 517-44. 2011.

Syabryaj, S., Molchanoff, S., Utescher,T. and Bruch, A.A.: Changes of climate and vegetation during the Miocene on the territory of Ukraine. Palaeogeogr., Palaeoclim., Palaeoecol., 253, 153-168, 2007.

Thonicke, K., Venevsky, S., Sitch, S., and Cramer, W.: The role of fire disturbance for global vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model. Global Ecology and Biogeography, 10, 661-677, 2001.

Utescher, T., Erdei, B., Francois, L., and Mosbrugger, V.: Tree diversity in the Miocene forests of Western Eurasia. Palaeogeogr., Palaeoclim., Palaeoecol., 253, 242-266, 2007.

Utescher, T., Bruch, A.A., Micheels, A., Mosbrugger, V., and Popova, S.: Cenozoic climate gradients in Eurasia—a palaeo-perspective on future climate change? Palaeogeogr., Palaeoclim., Palaeoecol., 304, 351–358, 2011.

Williams, M., Haywood, A.M., Taylor, S.P., Valdes, P.J., Sellwood, B.W., and Hillenbrand, C.D.: Evaluating the efficacy of planktonic foraminifer calcite delta 18 O data for sea surface temperature reconstruction for the Late Miocene. Geobios, 38, 843–863, 2005.

Wolfe, J.A.: Tertiary climatic changes at middle latitudes of western North America. Palaeogeogr., Palaeoclim., Palaeoecol., 108, 195–205, 1994a.

Wolfe, J.A:. An analysis of Neogene climates in Beringia. Palaeogeogr., Palaeoclim., Palaeoecol., 108, 207–216, 1994b.

Zaehle, S., Sitch, S., Smith, B. & Hatterman, F. 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochemical Cycles 19: 3020.

Zaehle, S., Medlyn, B.E., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hickler, T., Luo, Y., Wang, Y.-P., El-Masri, B., Thornton, P., Jain, A., Wang, S., Warlind, D., Weng, E., Parton, W., Iversen, C.M., Gallet-Budynek, A., McCarthy, H., Finzi, A., Hanson, P.J., Prentice, I.C., Oren, R. & Norby, R.J. 2014. Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO_2 Enrichment studies. New Phytologist 202: 803–822.

Zhang Y.G., Pagani M., Liu Z., Bohaty S.M. and DeConto R.: A 40-million-year history of atmospheric CO_2. Philosophical Transactions of the Royal Society A, 371, 20130096, 2013.

2143 Tables

2144

2145 Table 1

		MODEL			
DATA		Absent	Trace	Sub-dominant	Dominant
	Absent	0	0	-1	-2
	Trace	0	1	0	-1
	Sub-dominant	-1	0	1	0
	Dominant	-2	-1	0	2

2146

2147 Table 1: Contributions to the Agreement Index for each combination of data and

2148 model statuses.

2149

2150 Table 2

2151

	$CO_{2,clim}$ = 280 ppm		$CO_{2,clim}$ = 450 ppm		
Region	$CO_{2,veg}$ = 280 ppm	$CO_{2,veg}$ = 450 ppm	$CO_{2,veg}$ = 280 ppm	$CO_{2,veg}$ = 450 ppm	Number of fossil sites
Global	-0.67	-0.6	-1.02	-0.96	-0.96
Europe	0.01	0.04	-0.22	-0.23	103
(Central Europe)	(0.2)	(0.19)	(-0.01)	(-0.04)	(57)
Asia	-0.46	-0.44	-0.58	-0.54	37
North America	-0.1	-0.07	-0.05	-0.07	19
Central and South America	-0.04	-0.07	-0.04	-0.05	3
Africa	-0.05	-0.02	-0.07	-0.05	3
Australia	-0.03	-0.04	-0.04	-0.02	2

2152

2153

2154 Table 2: Global and regional Agreement Index values from all permutations of 280

2155 ppm and 450 ppm CO_2 concentrations in the climate model ($CO_{2,clim}$) and vegetation

2156 model ($CO_{2,veg}$) ~~models~~. Central Europe is shown separately and is defined to lie in

2157 the longitude range [0°, 25°] and latitude range [45°, 55°].

Table 3

		Vegetation CO_2	
		280 ppm	450 ppm
Climate CO_2	280 ppm	0.17	0.19
	450 ppm	0.01	0.03

Table 3: Central European Agreement Index values from all permutations of 280 ppm and 450 ppm CO2 concentrations in the climate and vegetation models. For these purposes, Central Europe is defined to lie in the longitude range [0, 25] and latitude range [45, 50].

2166 Figure captions

2167

2168 Figure 1. Modelled Late Miocene (Tortonian, 7-11 Ma) vegetation, using the
2169 ECHAM5-MPIOM AOGCM to drive LPJ-GUESS. A) The biome distribution with
2170 280 ppm CO_2 concentration, with the Agreement Index (AI) match overlain for
2171 palaeobotanical data. B) The biome distribution with 450 ppm CO_2 concentration,
2172 with the AI match overlain for palaeobotanical data. C) The dominant PFTs, with
2173 palaeobotanical data classified with same PFT scheme as the model overlain, with
2174 280 ppm CO_2 concentration. D) The dominant PFTs, with palaeobotanical data
2175 classified with same PFT scheme as the model overlain, with 450 ppm CO_2
2176 concentration.

2177

2178 Figure 2. Agreement Index with the null model distribution and the AI values shown
2179 for model runs with different CO_2 concentration.

2180

2181 Figure 3. Modelled grass fraction of Leaf Area Index (LAI) for present-day
2182 simulation, Tortonian 280 ppm CO_2, and Tortonian 450 ppm CO_2 concentrations,
2183 respectively. Shown also is the grass fraction of LAI for a mixed CO_2 forcing in
2184 climate and vegetation model.

2185

2186 Figure 4. Modelled tree fraction of Leaf Area Index (LAI) for present-day simulation,
2187 Tortonian 280 ppm CO_2, and Tortonian 450 ppm CO_2 concentrations, respectively.
2188 Shown also is the tree fraction of LAI for a mixed CO_2 forcing in climate and
2189 vegetation model.

2190

Figure 5. Agreement Index difference between the 280 ppm and 450 ppm runs.

Figures

Figure 1

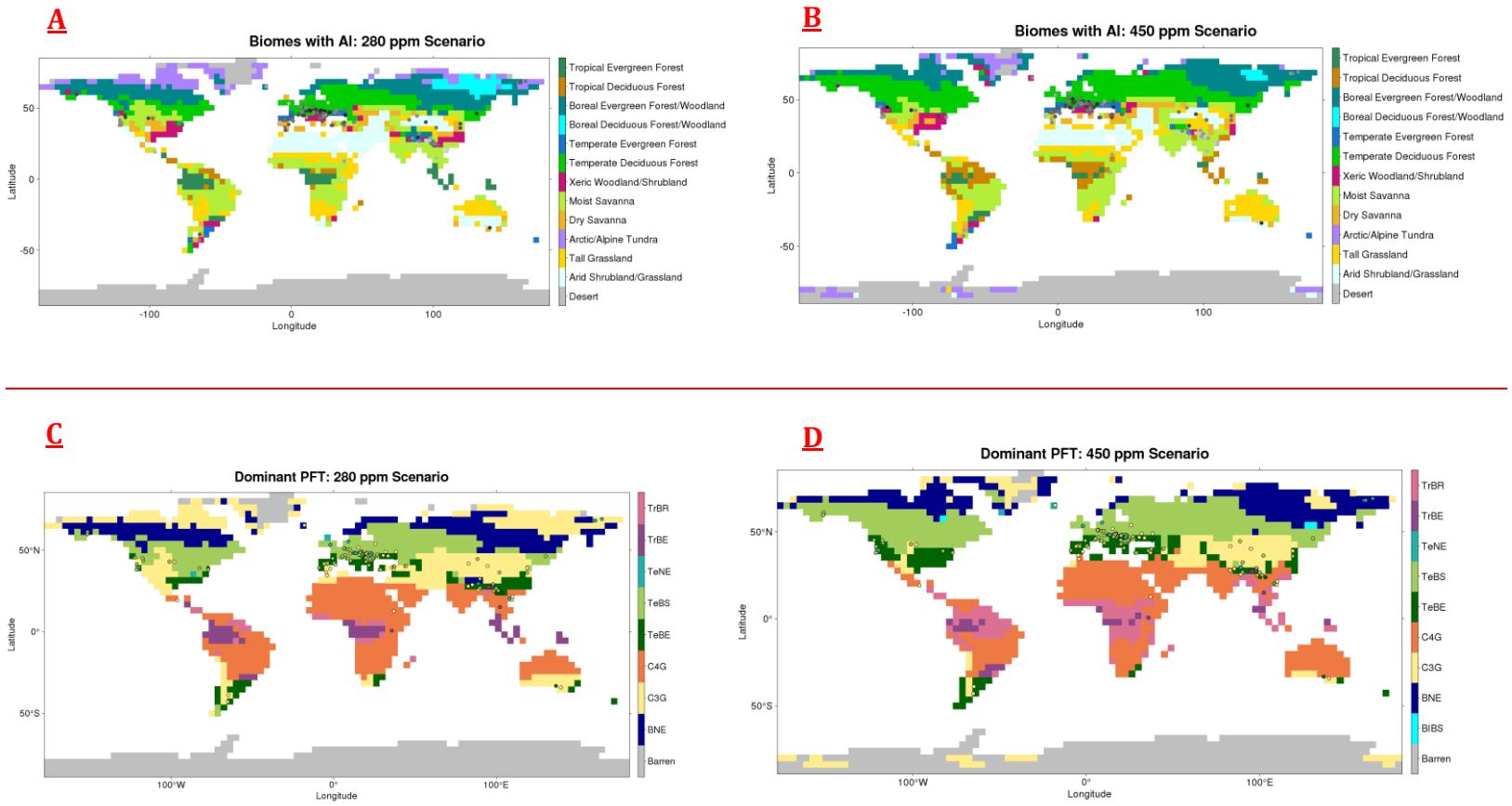


Figure 2

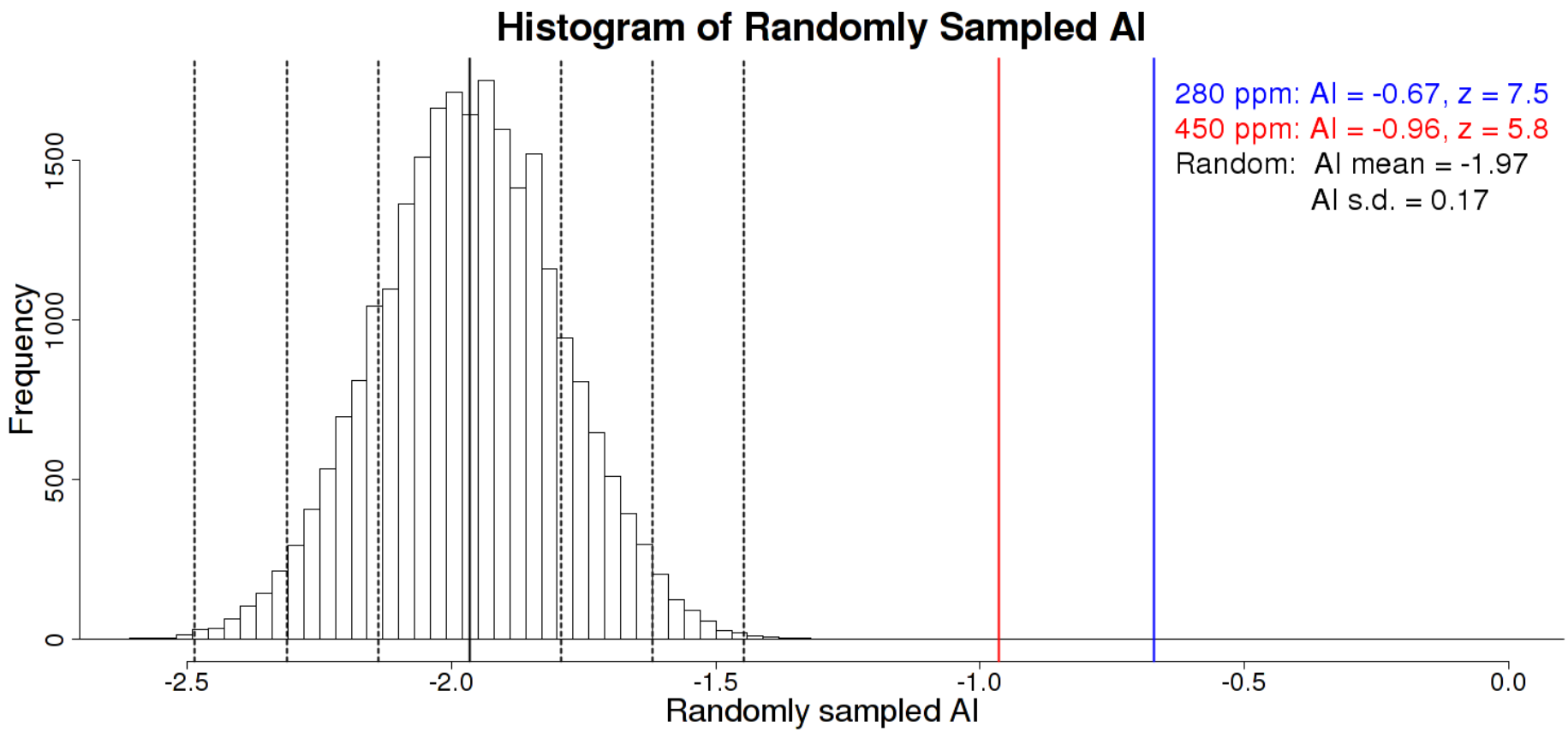

Histogram of Randomly Sampled AI
280 ppm: AI = -0.67, z = 7.5
450 ppm: AI = -0.96, z = 5.8
Random: AI mean = -1.97
AI s.d. = 0.17
Frequency
0
500
1000
1500
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
Randomly sampled AI

Figure 3

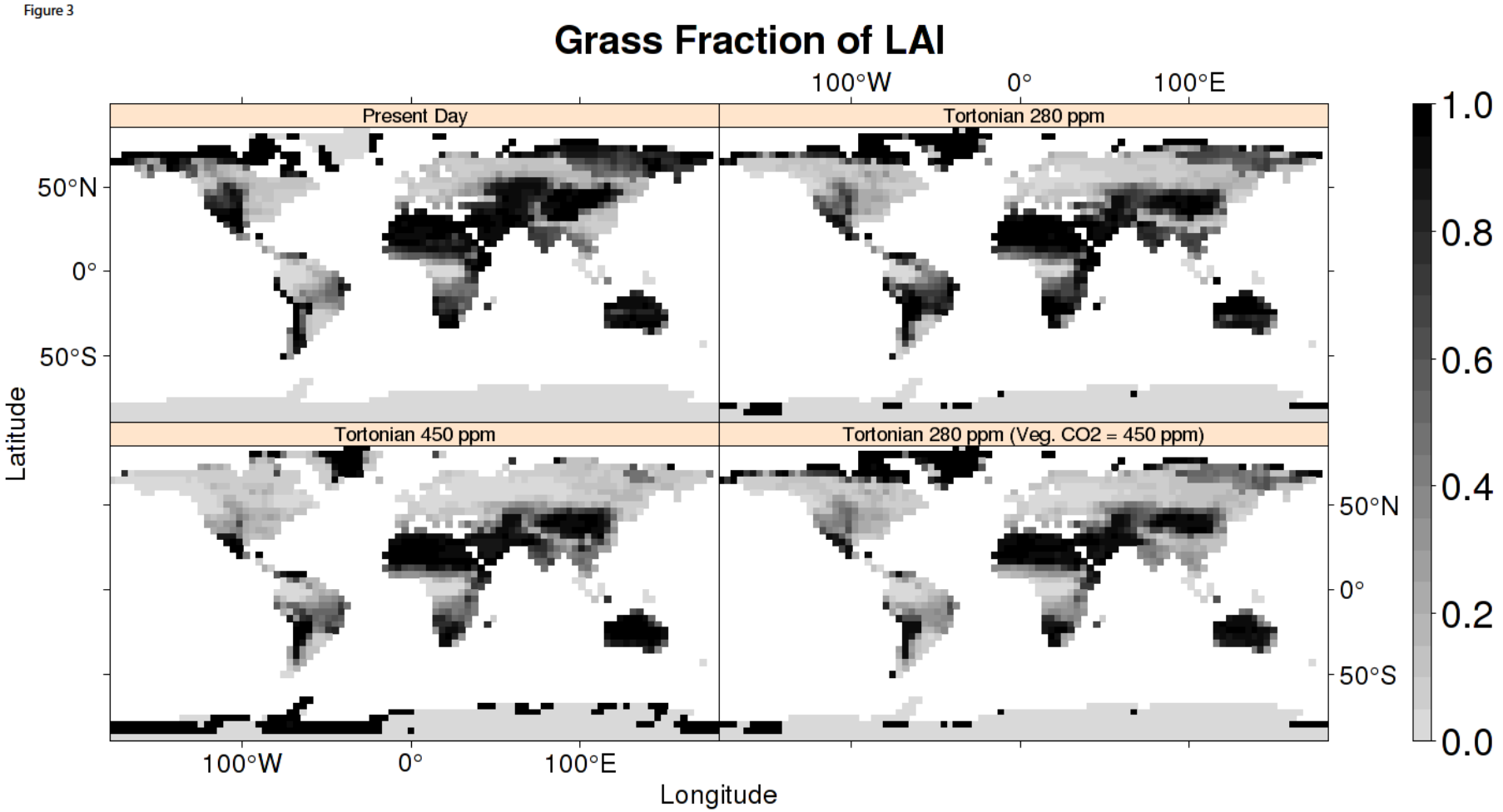


Figure 3
Grass Fraction of LAI
Present Day
Tortonian 280 ppm
Tortonian 450 ppm
Tortonian 280 ppm (Veg. CO2 = 450 ppm)
100°W
0°
100°E
50°N
0°
50°S
Latitude
Longitude
1.0
0.8
0.6
0.4
0.2
0.0

Figure 4

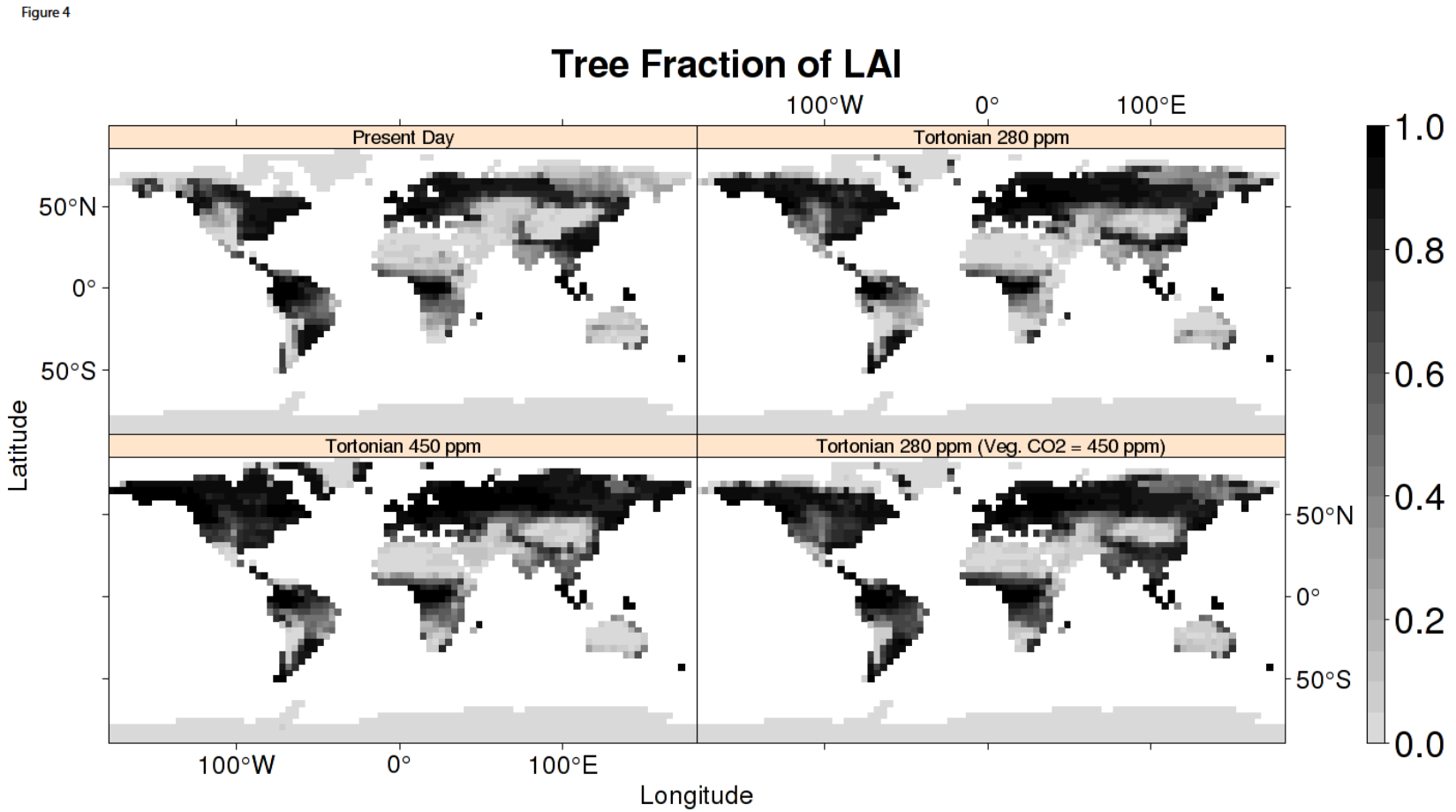
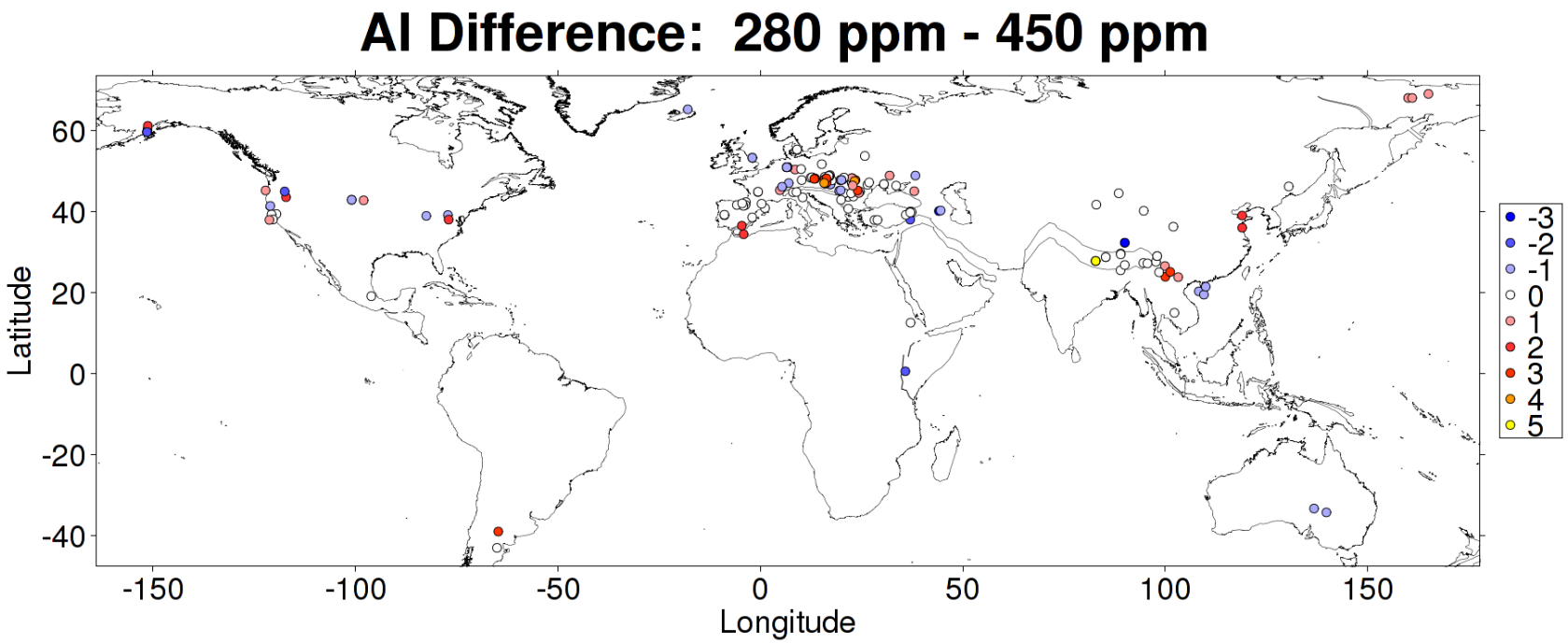


Figure 4
Tree Fraction of LAI
Present Day
Tortonian 280 ppm
Tortonian 450 ppm
Tortonian 280 ppm (Veg. CO2 = 450 ppm)
100°W
0°
100°E
50°N
0°
50°S
Latitude
Longitude
1.0
0.8
0.6
0.4
0.2
0.0

Figure 5


AI Difference: 280 ppm - 450 ppm
Latitude
60
40
20
0
-20
-40
-150
-100
-50
0
50
100
150
Longitude
-3
-2
-1
0
1
2
3
4
5

Appendices

Appendix A: Plant Functional Types (PFTs)

The ~~used~~ PFTs used here follow from Ahlström et al. (2012) with some modifications as noted in the main text. In particular, the parameters for shade-tolerance classes, leaf forms, and growth types are unchanged from Ahlström et al. (2012, Table S2). Table A1 gives a complete list of the PFTs and their parameters, as used in this study.

Appendix B: Biome classification~~.~~

The biome classification used here is shown in Table B1. ~~is based on the classification used in Hickler et al. (2006) but includes the modifications used in~~It is almost identical to that of Smith et al. (2014) but~~. It is further~~ slightly modified because the shade intolerant broad-leaved summergreen (IBS) PFT in Smith et al. (2014) has been split into a temperate shade intolerant broad-leaved summergreen (TeIBS) PFT and a boreal shade intolerant broad-leaved summergreen (BIBS) PFT for this study. In this classification BIBS is treated as IBS for classifying boreal forests, and TeIBS is added to TeBS when classifying temperature forests. Furthermore, to classify alpine tundra as well as arctic tundra, tundra is mapped if GDD_5 < 400 °C·days (GDD_5 = annual accumulated degree-day sum of days above 5°C)

Appendix Tables

Table A1 PFT Specific Parameters

PFT	Phenology	Shade tolerance class	Leaf Type	Growth Form	$T_{c,\,min}$ (°C)	$T_{c,\,max}$ (°C)	GDD_5 (°C day)	r_{fire}	a_{leaf} (year)	A_{ind} (year)	$Tr\text{-}_{leaf}$ (year^{-1})	Br (gC gN^{-1} day^{-1})	T_{opt} (°C)
BNE	evergreen	tolerant	needle-leaved	tree	-32.5	-2	600	0.3	3	500	0.33	2	10-25
BINE	evergreen	intolerant	needle-leaved	tree	-32.5	-2	600	0.3	3	500	0.33	2	10-25
BNS	deciduous	intolerant	needle-leaved	tree	-	-2	350	0.3	0.5	300	1	2	10-25
BIBS	deciduous	intolerant	broad-leaved	tree	-	-2	350	0.1	0.5	200	1	2	10-25
TeBS	deciduous	tolerant	broad-leaved	tree	-17	15.5	1200	0.1	0.5	400	1	1	15-25
TeIBS	deciduous	intolerant	broad-leaved	tree	-17	15.5	1200	0.1	0.5	200	1	1	15-25
TeBE	evergreen	tolerant	broad-leaved	tree	3	18.8	1200	0.3	3	300	0.33	1	15-25
TeNE	evergreen	intolerant	needle-leaved	tree	-2	22	900	0.3	3	300	0.33	1	15-25
TrBE	evergreen	tolerant	broad-leaved	tree	15.5	-	-	0.1	2	500	0.5	0.15	25-30
TrIBE	evergreen	intolerant	broad-leaved	tree	15.5	-	-	0.1	2	200	0.5	0.15	25-30
TrBR	deciduous	intolerant	broad-leaved	tree	15.5	-	-	0.3	0.5	400	0.5	0.15	25-30
C3G	-	-	-	grass	-	-	-	0.5	0.5	-	1	1	10-30
C4G	-	-	-	grass	15.5	-	-	0.5	0.5	-	1	0.15	20-40

Table A1. PFT characteristics and parameter values used in this study. $T_{c,min}$ = Minimum coldest-month temperature for survival and establishment; $T_{c,max}$ = maximum coldest-month temperature for establishment; GDD_5 = Minimum accumulated degree-day sum of days above 5°C for establishment; r_{fire} = Fraction of individuals surviving fire; a_{leaf} = leaf longevity; a_{ind} = individual maximum, non-stressed longevity; Tr_{leaf} = Leaf turnover rate; Br = Base respiration rate at 10°C; T_{opt} = Optimal temperature range for photosynthesis. Full PFT names: BNE = boreal needle-leaved evergreen tree; BINE = boreal shade intolerant needle-leaved evergreen tree; BNS = boreal needle-leaved summergreen tree; BIBS = boreal shade intolerant broad-leaved summergreen tree; TeBS = temperate broad-leaved summergreen tree; TeIBS = temperate shade intolerant broad-leaved summergreen tree; TeBE = temperate broad-leaved evergreen tree; TeNE = temperate needle-leaved evergreen tree; TrBE = tropical broad-leaved evergreen tree; TrIBE = tropical shade intolerant broad-leaved evergreen tree; TrBR = tropical broad-leaved raingreen tree; C3G = C_3 grass; C4G = C_4 grass.

Table B1 Biome classification scheme for model output

Biome[13]	Tree LAI[1]	Grass LAI[1]	Total LAI[1]	Domiant Tree PFT[2]
Tropical rainforest[6]	> 2.5			TrBE[3]
Tropical deciduous forest[7]	> 2.5			TrBR
Tropical seasonal forest[8]				TrBE[3] or TrBR
Boreal evergreen forest/woodland[9]	> 0.5			BNE[4] or BIBS
Boreal deciduous forest/woodland[9]	> 0.5			BNS
Temperate broadleaved evergreen forest[10]	> 2.5			TeBE
Temperate deciduous forest[10]	> 2.5			TeBS[5]
Temperate/boreal[11] mixed forest	> 2.5			
Temperate mixed forest				
Xeric Woodlands/ Shrublands	0.5-2.5	< 20% of total		
Moist Savnna	0.5-2.5		> 2.5	
Dry Savanna	0.5-2.5		≤ 2.5	
Arctic/alpine tundra[12]	< 0.5		> 0.2	
Tall grassland		> 2.0		
Arid shrubland/ steppe (1)	> 0.2	< 1.0		
Dry grassland		> 0.2		
Arid shrubland/ steppe (2)			> 0.2	
Desert			≤ 0.2	

[1] Growing season maximum leaf area index; [2] Highest LAI; PFTs are listed in Table A1, [3] TrBE + TrIBE, [4] BNE + BINE, [5] TeBS + TeIBS, [6] Mapped if $LAI_{TrBE} > 0.5 \cdot LAI_{trees}$; [7] Mapped if $LAI_{TrBR} > 0.5 \cdot LAI_{trees}$; [8] Mapped if $LAI_{tropical\ trees} > 0.5 \cdot LAI_{trees}$ and TrBE or TrBR has highest LAI among trees; [9] Mapped if $LAI_{boreal\ trees} > 0.5 \cdot LAI_{trees}$; [10] Mapped if LAI_{TeBS} or $LAI_{TeBE} > 0.5 \cdot LAI_{trees}$; [11] Mapped if $0.2 \cdot LAI_{trees} < LAI_{boreal\ trees} < 0.8 \cdot LAI_{trees}$ and $0.2 \cdot LAI_{trees} < LAI_{temperate\ trees} < 0.8 \cdot LAI_{trees}$; [12] Mapped at latitude > 54° or GDD_5 (see Table A1 for definition) < 400°C·days; [13] Classification must be done in the same order as table.

Table B1 Classification scheme for deriving vegetation biomes from PFT abundances (leaf area index, LAI), following Smith et al. 2014.