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.

1. Introduction

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

units MN9 to MN12, and Vallesian and lower Turolian mammal zones (Steininger 1999). The boundary conditions used for the climate model, as well as the proxy data we use, are dated within these time slices. From here on, we just use the term Tortonian to indicate this time period, and refer to the Late Miocene when we discuss trends in more general terms.

Here we run the dynamic global vegetation model (DGVM) LPJ GUESS (Smith et al., 2001, Sitch et al., 2003, Ahlström et al., 2012) for the Tortonian with two different CO_2 concentrations to investigate the vegetation dynamics during this period. We use climate data simulated for the Tortonian by Knorr et al. (2011) and Knorr and Lohmann (2014), using a fully coupled AOGCM without any flux corrections. We concentrate on whether the DGVM can create and maintain the mid-latitude seasonal vegetation cover in a generally warmer world, as suggested by the proxy data, and on the sensitivity of the vegetation to CO_2 concentration. We compare our results with existing terrestrial proxy data and previous modelling results, and discuss the implications from our results. Our hypothesis is that in order to maintain the seasonal and open vegetation of the Late Miocene, we need low atmospheric CO_2 concentration.

2. Previous model studies

Several vegetation model runs have been performed previously for the Late Miocene period. One of the first was a BIOME4 model (Kaplan, 2001) run for the Tortonian by Micheels (2003) to interpolate between the vegetation reconstructed by qualitative 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. 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). 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), 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 for the present day to a present-day biome map. These results are

presented in Appendix C, where an examination of the model setup's ability to distinguish between present day and Tortonian vegetation can also be found.

3.4 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 further problem is that it is difficult to know how PFT diversities in the fossil record correlate to an abundance measure that can be simulated by a vegetation model. An example of a commonly used abundance measure from vegetation models is Leaf Area Index (LAI), that is the leaf area per unit ground area. Standard statistical tests, such as Spearmans's rank correlation and Pearson's production moment correlation coefficient, between modelled PFT LAI fraction and the PFT diversities in the fossil

record, did not yield useful results, possibly for the reasons discussed above. These results are shown and discussed in Appendix D.

3.4.1 Discussion of previous quantitative approaches

To go beyond simple visual comparisons of model and data, and for hypothesis testing, we require a quantitative measure of agreement between fossil data and model output. Different approaches have been developed to compare fossil data to model results with some quantitative element. The study of Pound et al. (2011) uses Cohen's kappa to determine biome agreement, comparing both the 27 "native" biomes from BIOME4 and a 7 "megabiome" classification. This does offers a single statistic which could be used for hypothesis testing. However, there are inherent shortcomings when using kappa to compare biome classifications and with biome classifications themselves.

The inherent disadvantage of comparing kappa scores for biomes is that kappa does not include any mechanism to account for "degrees of difference" 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 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 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.

3.4.2 Calculation of Agreement 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. Based on these fractions, each PFT is assigned one of 4 statuses 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 Appendix E. 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.

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 Appenix E. 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 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 F1 in Appendix F). 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 F2 in Appendix F). 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, 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 Sect. 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 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 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 facts strongly suggest 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 Appendix C.

The result that 280 ppm run agrees better with the palaeobotanical data poses a question: how can we have the combination 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, we see that with 280 ppm in the climate simulation there are more open conditions in North America, regardless of the 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. In other regions 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 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 2, Figure 5). 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. These conclusions are also supported by fossil mammal data (e.g. Fortelius et al., 2014).

In the 280 ppm run 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. 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, Pound et al. (2011) BIOME4 simulation produced 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 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 better than the 450 ppm simulation.

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.

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, with no difference between the different CO_2 simulations. The 280 ppm run shows more temperate broadleaved evergreen trees in southern and eastern China and the surrounding area, than in the 450ppm 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.. 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 S4a). 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 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.

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. There are still uncertainties in the models, and these results should be tested with different models. 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 suggest that atmospheric CO_2 levels were relatively low during the Late Miocene, and that the Late Miocene fossil vegetation data can be used in conjunction with vegetation/climate modeling to constrain CO_2 concentrations in the atmosphere.

Appendices

Appendix A: Plant Functional Types (PFTs)

The 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, their 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. It is almost identical to that of Smith et al. (2014) but 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 C Model benchmarking and effect size

Figure C1(a) compares the biome distributions from the present day PGF(Princeton Group Forcing, Sheffield et al., 2006) control run and potential natural vegetation biomes from Hickler et al (2006, modified from Haxeltine and Prentice, 1996), using

the biomes classification described in Appendix B. Figure C1(b) shows the dominant PFT. The simulation captures the broad patterns of present day vegetation. The reader is referred to Smith et al. (2014, their Figure 2(C)) for a more detailed qualitative comparison of the biomes deriving from LPJ-GUESS without the modifications employed for this study.

As noted in the main text, there is uncertainty in potential natural vegetation as different reconstructed biome maps can differ considerably (compare, for example, Haxeltine and Prentice (1996), Ramankutty and Foley (1999), Freidl et al. (2010), Olson et al. (2001)).There are also uncertainties when assigning biomes from model output due to the necessary use of arbitrary thresholds to define cut-offs between biomes. To mitigate these uncertainties and allow a meaningful quantitative comparison (Cohen's Kappa statistic), we follow the approach of Harrison and Prentice (2003) and Pound et al. (2011) and aggregate biomes to eight megabiomes. The biome aggregation is described in Table C1 and follows the scheme Harrison and Prentice (2003) with minor alterations. The megabiomes resulting from the aggregation are shown in Fig. C1(c). Calculating Cohen's Kappa between the data and model gives a value of 0.62, classified as "good" agreement by Monserud and

Leemans (1992)). We interpret this as sufficiently good agreement and therefore sufficient model skill for the purposes of this study.

To examine the model setup's overall sensitivity to CO_2 concentration and its ability to differentiate between present day and Tortonian climate, we calculated Cohen's Kappa between the simulated megabiome distributions. These comparisons only involve modelled biomes, and these modelled biomes are produced using identical classification schemes, so the concern raised above (and in Sect. 3.4.1 of the main text) about the uncertainty in biome classifications does not apply here. The issue of "degrees of difference" is still relevant, but is ameliorated to some extent by the use of the coarser megabiome scheme. The Kappa between the 280 ppm CO_2 and 450 ppm CO_2 reconstructions is 0.70. Given that the model setup is identical except for the CO_2 concentration and that all other factors are equal, we believe that this indicates a sufficiently large sensitivity to atmospheric CO_2 concentrations for the purpose of this study. The Kappa between the Tortonian 280 ppm biomes and the PGF control run biomes is 0.64, and comparison of the Tortonian 450ppm biomes and the PGF control run biomes gives a Kappa of 0.48. Considering again that these maps are produced with identical methodologies, these Kappa scores indicate that the method can well-distinguish between Tortonian vegetation and present day vegetation.

Appendix D Pearson's product moment correlation coefficients and Spearman's rank correlation coefficients

Both Pearson's product moment correlation coefficients and Spearman's rank correlation coefficients were calculated for the 280 ppm and 450 ppm scenarios per PFT and for the entire dataset. These are presented here in Fig. D1. As mentioned in the main 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 paleobotanical data, but is far from conclusive.

Appendix E Agreement Index robustness checks

The robustness of the AI with respect to the various subjective choices was tested as described below.

E.1 Choice of fractional ranges to define AI statuses

A factorial study was carried out with the following values for the fraction ranges.

Min for trace:	0.025, 0.05, 0.075 (original was 0.05)
Min for sub-dominant:	0.075, 0.15, 0.3 (original was 0.15)
Min for dominant:	0.5, 0.75 (original was 0.5)

The results are shown for the 450 ppm run versus the 280 ppm in Fig. E1. The default boundaries are marked with a red star. Overall, it is clear that the 280 ppm gives better

agreement than the 450 ppm in almost all cases. The exception (large black square) has a huge sub-dominant range from 0.075 to 0.75 which will include many PFTs, and therefore this combination of ranges has very little differentiating power.

The boundaries control the absolute value of the AI much more than they control the difference between the 280/450 ppm runs, which suggests that the scientific result are robust against changes in the boundaries. It is possible to choose different boundaries to get either better differentiating power or higher values (in terms of absolute numbers) or even both, but this study was performed as an a posteriori check of robustness, not to tune the method, so the initial choices were maintained.

E.2 Choice of numbers for the quantification of the different types of agreement

Table E1 shows the AI scores and ranges when different numbers are used to quantify agreement/disagreement between statuses. In all cases the score is higher for the 280 ppm run than for the 450 ppm run.

E.3 Estimation of random agreement

As discussed in Sect. 3.4.3 of the main text, there is no obvious method for simulating ‘random agreement’ to estimate agreement by chance. Simply assigning each PFT a random fraction (or AI status) will result in unrealistic PFT combinations and unrealistic proportions of absent vs. present PFTs which has a strong effect on AI

scores (since by construction of the method, absent PFT do not contribute to the AI score, they only reduce it when they are incorrectly simulated). The structure of the fossil data could be used to varying degrees when generating data to simulate random chance, but following this structure too closely could lead to artificially high levels of agreement chance as the supposedly random data are restricted to be very similar to the fossil data.

Here we define, test, and discuss models to estimate chance agreement and define four classes of model.

A. Models which use only the bare minimum of information from the fossil dataset. Specifically, the number of PFTs and the number of sites are important for assessing variability and so must be included. Apart from that, no further information from the fossil data is used. As such, these models rely mostly on the inherent properties of the AI method but are naive to most of the details of the data – let us call them 'naive methods'. In such methods both fossil data and model data are randomly generated.

B. Models which also use the structure of the fossil data, for example the distribution or mean number of non-absent PFTs per site or the distribution of PFT fractions, but not the fossil data themselves. From such structural information, both random fossil and model datasets are generated to mimic the structure of the fossil data. Let us call these "data-structured methods".

C. Models which use the fossil data directly and compare it to randomly generated model data. The randomly generated model data may or may not be informed by the fossil data (as in data structured methods). Let us call the methods

"data-centered methods".

D. Models which compare fossil data to randomly sampled model data output. These methods have the advantage that randomly sampled model data is guaranteed to be ecologically sensible (insofar as the vegetation model is sensible). Let us call these "model-sampled methods".

Examining the fossil data shows that the mean number of non-absent PFTs per fossil site is 4.2 (4 used when an integer number is required when constructing the models below), with the distribution shown in Fig. E2(a). This simple distribution is simulated exactly when building the chance agreement models B2, B4, C2 and C4, as described below. The distribution of PFT fractions across all sites and PFTs is shown in Fig. E2(b). This can be well approximated by simulating each PFT abundance/diversity as the exponential of a random number drawn from a Gaussian distribution with mean = 1.0 and standard deviation = 1.75, and then calculating PFT fractions by dividing by the total abundance/diversity at the site (exactly as one would do to calculate PFT fractions from abundance/diversity data). This formulation was found by trial-and-error, but as can be seen in Fig. E2(b), it matches the fossil data extremely well. In particular the first bin (which marks the 0.05 cut-off below which a PFT is considered absent) is extremely well simulated.

We present the mean and standard deviation for a range of chance agreement methods (each category is represented) and compare the resulting *Z*-scores and *p*-values for the 280 ppm and 450 ppm simulations in Table E2. Each method has been employed with 5000 iterations (each iteration sums AI scores across all sites in the fossil dataset) and

the resulting distributions of AI scores are all consistent with a Gaussian distribution by visual inspection, and by inspection of a quantile-quantile (QQ) plot (data not shown), as would be expected by the Central Limit Theorem. The models are:

A. Naive models

Model A1: Both model and data are generated such that each PFT is assigned a fraction with equal probability. The fractions are then normalised to sum to unity.

Model A2: Both model and data are generated such that each PFT is assigned an AI status with equal likelihood, with the addition restriction that only one dominant PFT can be assigned per site.

B. Data-structured models

Model B1: Both model and data are generated such that 4 PFTs are assigned a non-absent AI status with equal likelihood (the rest are assigned absent), with the addition restriction that only one dominant PFT can be assigned per site.

Model B2: Both model and data are generated such that a random number of PFTs are assigned a non-absent AI status with equal likelihood (the rest are assigned absent), with the random number chosen from a distribution which matches the fossil data, and the additional restriction that only one dominant PFT can be assigned per site.

Model B3: Both model and data are generated such that 4 PFTs are assigned a non-zero fraction with equal probability. The fractions are then normalised to sum to unity.

Model B4: Both model and data are generated such that a random number of PFTs is assigned a non-zero fraction with equal probability, with the random number chosen from a distribution which matches the fossil data. The fractions are then normalised to sum to unity.

Model B5: Both model and data are generated such that the PFT fractions have the same distribution as the fossil data (as described above).

C. Data-centered models

Models C1-C5 are the same as models B1-B5 except that the fossil data are not simulated, instead the actual fossil data are used. In other words, models B1-B5 are data-structured models, and models C1-C5 are the data-centered analogs.

Models C6 and C7 are the same as models A1 and A2, except that the fossil data is not simulated; instead, the actual fossil data is used. In other words, models C6-C7 are the data-centered analogs of naive models A1 and A2.

D. Model-sampled models

Model D1: The real fossil data are used and each fossil site is matched to a randomly determined grid cell from either the 280 ppm or 450 ppm simulations. This is the model presented in the main text.

Model D2: The real fossil data are used and each fossil site is matched to a randomly determined grid cell from either the 280 ppm or 450 ppm simulations, with the additional restriction that the modelled grid cell must be in a latitude band of +/- 10 degrees around the fossil site (corresponding to approximately 3 grid boxes on either side), or in the mirror image latitude band in the other hemisphere.

Examining the Table E2, we see that the naive models (A1 and A2) produce a relatively high estimation of agreement by chance. In fact, quantifying agreement by chance using model A1 gives such a high level of agreement that negative *Z*-scores for the 280 ppm and 450 ppm runs are produced. However, this level of agreement is unrealistic. This is because these models make no assumptions about the structure of

the fossil data, so must necessarily assume a rather homogeneous structure, with fractions (in model A1) and status (in model A2) having equal likelihood (except for the dominant status in A2, which can be restricted to one per site). This homogeneous data structure produces relatively high degree of agreement by chance. If one (non-absent) category is produced very often for PFTs in both the simulated model data and the simulated fossil data, there will be a high chance of a match, and therefore a high AI score. This is particularly pronounced in the model A1, which produces many more non-absent PFTs in the randomly generated data than are seen in the data. In particular, high numbers of trace statuses are produced because in model A1 each fraction has an expectation value of $1/N$, where N is the total of PFTs compared, in this case 10. This gives an expectation value of 0.1, which is right in the middle of the fractional range for trace status. Comparing the fractions of each status produced: model A1 produces the following percentages of classifications: 24/55/21/0% (absent/trace/subdominant/dominant), whereas the fossil record shows 58/21/16/5%. These highly disparate percentages show that this method of generating data produces datasets which are very different from the fossil data used, so it is not a meaningful estimate of agreement by chance in the context of this analysis. This conclusion is further reinforced by the results of model C7, which is the equivalent data-centered model to the naive model. This model, which compares data generated by model A1 with real fossil data, shows much lower agreement than model A1, indicating that the method of simulating data does not match well the real fossil data.

Model A2 shows a much lower level of agreement by chance than model A1. This is because absent, trace and subdominant statuses are produced with equal probability, so, unlike model A1, the trace classification is not overwhelming. Without the

tendency for one status to be produced in such large quantities, the simulated data are less homogeneous and therefore estimate less agreement by chance. This gives a more reasonable estimate of agreement by chance. With this model, the *p*-values for getting better agreement from randomly generated data are estimated to be $p < 0.05$ for the 450 ppm scenario, and $p < 10^{-4}$ for the 280 ppm scenario. It should be noted that this model still does not produce data with a similar structure to the fossil data (30/30/30/9% absent/trace/subdominant/dominant compared 58/21/16/5%, note in particular the under representation of absence), so it is not a particularly good estimation of agreement by chance.

The data-structured and data-centered models all produce much less agreement by chance than the naive models. This is reasonable as these models use the structure of the fossil dataset to produce random data which are structured more like the fossil data, and this structure (as it is less homogeneous) decreases the agreement by chance. The *Z*-scores were very much higher, all greater than 10, corresponding to *p*-values which are so small that no meaningful comparison is possible. All that can be said is that the probability of getting better agreement by chance according to one of these chance agreement models is vanishingly small. Models C5 and B5 (which use simulated PFT fraction very similar to the actual fossil data and so mimic the real data most closely) give very similar results to model D1 (presented in the main text)

The final category, model-sampled models, estimates higher agreement by chance than the data-centered or data-structured models. They also have the desirable feature that only ecologically realistic PFT (according to the vegetation model) are produced. The more restrictive model of the two chance agreement models (model D2, which

requires the random modelled to be within 10 degrees latitude of the matching fossil site), gives *Z*-scores above 4.5 for the 280 ppm scenario and above 2.5 for the 450 ppm scenario. This gives a *p*-value for getting better agreement from randomly generated data to be $p < 10^{-2}$ for the 450 ppm scenario and $p < 10^{-5}$ for the 280 ppm scenario. The 'looser' model (model D1, presented in the main text) gives much higher *Z*-scores and extremely small *p*-values for both CO_2 scenarios.

To summarise, a selection of chance agreement models have been examined. All models which produce data with structure with some reasonable correspondence to the actually fossil data indicate that both the Tortonian vegetation simulations presented here agree better with the fossil data than simulated chance agreement by a considerable margin. Furthermore, the standard deviations of all models range between 0.08 and 0.33. Based on these values, the *Z*-score of the 280 ppm scenario shows better agreement than the 450 ppm simulation, by between 0.88 and 3.4 units of standard deviation. In 11 out of 16 models examined here, the difference was greater than 1.5 units of standard deviation. We believe this (and the other robustness check detailed above) demonstrates the robustness of the AI method and supports the scientific conclusions in the main text.

Appendix F: Details of paleobotanical data sites and classification

Table F1 lists the fossil sites used in this analysis, and Table F2 shows the classification from species or genera to the PFTs used in LPJ-GUESS.

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.

Friedl, M.A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley and X. Huang (2010), MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, 2001-2012, Collection 5.1 IGBP Land Cover, Boston University, Boston, MA, USA

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.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., ... & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience, 51(11), 933-938.

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.
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.

Tables

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

Table 1: Contributions to the Agreement Index for each combination of data and model statuses.

Table 2

	$CO_{2,clim}$ = 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

Table 2: Global and regional Agreement Index values from all permutations of 280 ppm and 450 ppm CO_2 concentrations in the climate model ($CO_{2,clim}$) and vegetation model ($CO_{2,veg}$). Central Europe is shown separately and is defined to lie in the longitude range [0°, 25°] and latitude range [45°, 55°].

Table A1

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_{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[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; [12] 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.

Table C1

Megabiome	Smith et al. (2014) biome (see Appendix B)
Tropical forest	Tropical seasonal forest Tropical evergreen forest
Temperate evergreen forest[1]	Temperate evergreen forest
Temperate deciduous forest[2]	Temperate conifer forest Temperate mixed forest Temperate/boreal mixed forest Temperate mixed forest
Boreal forest	Boreal deciduous forest/woodland Boreal evergreen forest/woodland
Savanna and dry woodlands	Xeric woodlands/shrub Moist savanna Tropical deciduous forest[3]
Grasslands and dry shrublands	Tall grassland Short grassland Dry savanna[4] Arid shrublands/steppe
Tundra[5]	Tundra
Desert	Desert

1 Denoted "warm temperate forest" in Harrison and Prentice (2003)
2 Denoted "temperate forest" in Harrison and Prentice (2003)
3 Tropical deciduous forest corresponds more closely to savanna types in Olson et al. (2001) and Friedl et al. (2010).
4 Dry savanna corresponds more closely to shrubland and grasslands types in Olson et al. (2001) and Friedl et al. (2010).
5 Only one tundra classification is distinguished here.

Table C1 Biome aggregation scheme following Harrison and Prentice (2003).

Table E1

	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 E1. 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.

Table E2

Model	Mean	S.D.	280 ppm *Z* score	450 ppm *Z* score	Z score difference (280 ppm - 450 ppm)	280 ppm *p*-value	450 ppm *p*-value
A1	2.48	0.17	-18.33	-20.02	1.69	1.00	1.00
A2	-3.43	0.17	15.97	14.29	1.68	0.0	0.0
B2	-3.35	0.17	15.51	13.83	1.68	0.0	0.0
B3	-6.24	0.32	17.30	16.40	0.90	0.0	0.0
B4	-6.26	0.33	16.92	16.04	0.88	0.0	0.0
B5	-2.23	0.15	10.10	8.22	1.88	0.0	0.0
C1	-2.97	0.15	15.29	13.36	1.92	0.0	0.0
C2	-2.94	0.15	15.16	13.23	1.93	0.0	0.0
C3	-5.72	0.29	17.51	16.51	1.01	0.0	0.0
C4	-5.70	0.29	17.35	16.35	1.00	0.0	0.0
C5	-2.31	0.14	11.64	9.59	2.06	0.0	0.0
C6	-4.74	0.11	36.03	33.46	2.57	0.0	0.00
C7	-1.94	0.09	14.78	11.40	3.38	0.0	0.0
D1	-1.96	0.17	7.51	5.83	1.69	2.86×10^{-14}	2.82×10^{-09}
D2	-1.35	0.15	4.66	2.69	1.98	1.54×10^{-06}	3.58×10^{-03}

Table E2 Mean value and standard deviation of chance agreement estimated from a selection of models, with 5000 full comparisons of data and model at all 167 fossil sites used in the analysis. Also shown are the Z-scores for the 280 ppm and 450 ppm vegetation reconstructions and the difference between them, and the *p*-values calculated from these Z-scores. A value of 0.0 in the p-value column implies $p < 10^{-15}$ or smaller.

Table F1

North America

Longitude	*Latitude*	*Region*	*Locality Code*
-151.5	59.6	North America	Lower Homerian AK
-151.5	59.6	North America	Middle Homerian AK
-151.5	59.6	North America	Upper Homerian AK
-151.4	60.2	North America	Lower Clamgulchian AK
-151.3	61.12	Cook Inlet Region, Alaska, USA	Chuitna River
-122.22	45.19	Northern America	Faraday
-121.27	37.93	North America	Neroly CA
-121.06	41.37	California, USA	Upper Cedarville Pit
-120.75	39.28	California, USA	Remington Hill
-120.38	38.03	California, USA	Table Mountain
-119.55	39.38	Nevada, USA	Chalk Hills
-117.5	44.95	Oregon, USA	Unity Or
-117.16	43.53	Eastern Oregon, USA	Succor Creek
-100.96	42.88	North America	Kilgore
-100.96	42.88	North America	Kilgore (pollen)
-98	42.75	Antelope County, Nebraska, USA	Antelope Ne
-96.11	19.12	Mexico	Paraje Solo Fm
-82.52	38.92	USA	Gray Sinkhole
-77.18	39.13	North America	Bryn Mawr
-77	38	South Maryland, USA	Brandywine Mar

South America

Longitude	*Latitude*	*Region*	*Locality Code*
-65.05	-42.94	Argentina	Puerto Madryn Fm
-64.74	-38.92	Argentina	Barranca Final Fm

Western Eurasia

Longitude	*Latitude*	*Region*	*Locality Code*
-17.939	65.187	Iceland	Fnjoskadalur Fm
-8.9	39.2	Portugal	Povoa 3
-8.87	39.06	Portugal	Azambuja
-5.8	41.6	Duero, Spain	Abezames
-4.589	36.491	Spain	Andalucia G1
-4.5	42	Duero, Spain	Torrem2
-4.2	41.4	Duero, Spain	Penafiel
-4.14	34.39	Marocco	Taza Guercif
-3.7	41.6	Duero, Spain	Burgos
-3.58	42.32	Spain	Castrillo del Val
-2.02	38.544	Spain	Rambla del Mojon 30 35
-2	53.25	Derbyshire, England	Derbyshire
-0.6	44.8	Landas, Spain	Arjuzanx
-0.57	44.87	France	Pont de Gail
0.3	41.9	Pirineo, Spain	Seo De Urgell
1.15	40.84	Tarragona, Spain	Tarragona E2 1
4.81	45.24	France	Andance
5.35	45.95	France	Amberieu S3
5.35	46.1	France	Soblay
6.47	50.92	Lower Rhine Basin, Germany	H7FB(F)
6.509	50.9	Lower Rhine Basin, Germany	H7F(B)
6.509	50.9	Lower Rhine Basin, Germany	H7F(F)

6.509	50.9	Lower Rhine Basin, Germany	H7FT(F)
6.691	50.954	Lower Rhine Basin, Germany	FO7(F)
6.691	50.954	Lower Rhine Basin, Germany	FO7O(B)
6.691	50.954	Lower Rhine Basin, Germany	FO7U(P)
6.71	50.91	Germany	FI7O(B)
7	47	Switzerland	Nebelberg
8.05	44.75	Piemonte, Italy	Guarene(F)
8.57	50.35	Mainz Basin, Germany	Dorheim (F)
8.9	44.8	Piemonte, Italy	Scrivia (F)
9.04	55.29	Denmark	Gram clay pit (J11)
10.05	50.45	Rhön Mountains, Germany	Wüstensachsen (F)
10.2	47.75	Southern Germany	Geissertobel(B)
10.43	43.48	Toscana, Italy	Gabbro(F)
12.4	48.3	Southern Germany	Aubenham (B)
12.75	48.45	Southern Germany	Lerch(B)
13.32	48.04	Austria	Schneegattern (B)
13.36	48.16	Austria	Grossenreith(B)
13.42	48.15	Austria	Lohnsburg(B)
13.55	48.1	Austria	Ampfelwang (F)
15.16	51.67	Southwest Poland	Godznica(F)
15.75	47.02	Steiermark, Kirchberg an der Raab, Austria	Wörth (B)
15.83	47.92	Burgenland, Austria	Neusiedl(B)
15.88	48.53	Vienna Basin, Austria	Ebersbrunn (B)
16	46.91	Austria	Mataschen rev Hably
16.08	46.93	Steiermark, Neuhaus/Klausenbach, Austria	Neuhaus(B)
16.08	46.93	Steiermark, Neuhaus/Klausenbach, Austria	Neuhaus rev Hably
16.27	48.17	Vienna Basin, Austria	Laaerberg(B)
16.33	48.17	Vienna Basin, Austria	Vösendorf(B)
16.36	47.15	Hungary	Sé (B)
16.364	48.023	Austria	Hennersdorf
16.58	48.03	Austria	Goetzendorf
16.88	48.75	Czech Republic	Postorna
16.88	48.75	Czech Republic	Postorna Moravska Nova Ves
17.05	48.7	Slovakia	Moravian Basin F(B)
17.05	48.7	Slovakia	Moravska N V(B)
17.17	48.97	Slovakia	Mistrin (B)
17.295	46.691	Hungary	Balatonszentgyorgi
17.635	47.684	Hungary	Gyor Sashegy
19.45	45.1	Serbia	Sremska
19.75	47.75	Hungary	Rozsaszentmarton (B)
19.75	47.75	Hungary	Rozsaszentmarton (rev. Hably)
19.84	45.23	Serbia	Sremska Kamenica
19.917	42.883	Montenegro	Popovici
20.032	47.776	Hungary	Visonta(B)
20.032	47.776	Hungary	Visonta rev Hably
20.4	47.97	Hungary	Felsötarkany
20.4	47.97	Hungary	Felsotarkany rev Hably
20.45	44.31	Serbia	Dubona I (B)
20.45	44.31	Serbia	Dubona II (B)
20.63	48.38	Hungary	Rudabanya (B)
20.75	44.52	Serbia	Durinci (B)
21.69	43.61	Serbia	Crveni Breg Grocka

21.71	40.68	Italy	Vegora
22.4	44.5	Serbia	Osojna
22.57	48.23	Carpathian area, Ukraine	Velikaya Began Pontian
22.58	46.97	Romania	Delureni (B)
22.67	48.23	Carpathian area, Ukraine	Velikaya Began N856well
22.8	46.4	Nagyfeketepatak, Bihor county, Romania	Valea Neagra(B)
22.983	43.7	Bulgaria	Drenovets Maeotian
23.25	47.5	Romania	Oas Basin
23.5	47.75	Romania	Chiuzbaia (rev. Hably)
24.02	45.18	Romania	Tanasesti Ramesti
24.32	44.57	Romania	Ramesti
24.6	44.9	Romania	Porceni
25.8	53.7	Belarus	Grodno Complex
26.44	46.58	Romania	Comanesti
26.86	47.17	Romania	Pau Iasi
28.2	37.9	Western Anatolia, Turkey	Nazilli Haskoy Upper Coal
28.925	37.92	Turkey, Western Anatolia	Saraykoy
30.52	46.75	Ukraine	Emetovka Early Maeotian 1
30.52	46.75	Ukraine	Emetovka Early Maeotian 2
31.91	48.86	Ukraine, western part, multiple sites	Western Ukraina (lower Maeotian)
33.53	46.37	Ukraine Plane, Ukraine	Chaplinka
35.93	39.17	Turkey	Sivas Karaozu
37	38	Central Anatolia, Turkey	Sivas Gemerek
37	40	Central Anatolia, Turkey	Duzyayla
37.018	39.754	Central Anatolia, Turkey	Sivas Vasiltepe
37.1	12.583	Ethiopia	Chilga
37.383	39.834	Central Anatolia, Turkey	Sivas Hafik
38	45	Western Georgia	Cocchati Complex
38.28	48.86	Ukraine, eastern part, multiple sites	Eastern Ukraina (lower Maeotian)
44.09	40.11	Armavir region, Armenia	Hoktemberya
44.53	40.24	Armenia	Hrazdan/2

Eastern Eurasia

Longitude	*Latitude*	*Region*	*Locality Code*
82.81	27.8	Nepal	Surai Khola 11-8 Ma
82.81	27.8	Nepal	Surai Khola 6-5 Ma
82.81	27.8	Nepal	Surai Khola 8-6 Ma
82.97	41.683	North Western China	Kuqa Xinjiang
85.3	28.75	China	Danzengzhukang Fm
85.3	28.75	China	Lower Woma Fm
88.5	44.5	North Western China	Southern Junggar Xinjiang
88.96	25.5	Bangladesh	Dupi Tila
89	29.43	China	Wulong
89	29.65	Tibet	Nanmulin Wulong Fm
90	26.8	Eastern Himalaya, Bhutan	Bhutan M, Siwalik
90	32.3	China	Lunpola Basin
90	32.3	Tibet	Lunpola Basin Dinquing 2
94.6	27.3	India	Assam Miocene
94.683	40.167	Northwestern China	Dunhuang
95.6	27.2	India	Deomali
97.7	27.6	India	Arunachal Pradesh
98	29	Tibet	Markam Lavula 1
98	29	Tibet	Markam Lavula a pollen

98.49	25.02	China	Tengchong
99.92	26.55	China	Jianchuan
100.017	23.9	China	Lincang
101.22	25.1	Southern China	Luehe Chuxiong
102	36.25	North Western China	Xining Minhe Basin
102.267	15.016	Thailand	Khorat
103.198	23.812	Yunnan, China	Xiaolongtan (Pre)
108.3	20.3	North continental shelf of South China Sea	Beibuwan 3
109.56	19.5	Coastal site South China Sea	Fushan depression Fushan 3
110	21.45	Coastal site South China Sea	Leizhou Peninsula Leizhou 3
119	36	Northern China	Bozhong Basin
119	39	Northern China	Bohai Gulf Basin
130.5	46.17	North Eastern China	Huanan Heilongjiang
136.75	-29.75	Australia	Stuart Creek
139.8	-30.7	Australia	Woltana1 Well 93,5
160	68	Siberia	Bayokov H1172
161	68	Siberia, Russia	Yanran H3690
165	69	Siberia, Russia	Nekkeiveem H3658 l mio

Africa

Longitude	*Latitude*	*Region*	*Locality Code*
35.8	0.6	Kenya	Tugen

Table F1 All palaeobotanical sites used in the study

Table F2

	PFTs	Main genera and species belonging to the PFTs
1	Tropical BL evergreen tree (TrBE)	*Abarema, Ehretia, Homalanthus, Litsea, Mastixia, Monotes, Moraceae, Ormosia, Phoebe, Polyspora, Sterculia, Tectocarya*
2	Tropical BL raingreen tree (TrBR)	*Acacia, Albizia, Cassia, Dalbergia, Dendropanax, Gleditsia*
3	Temperate NL evergreen tree (TeNE)	*Abies spp., Cathaya, Cedrus, Cephalotaxus, Keteleeria, Pinus spp., Podocarpus, Pseudotsuga, Sequoia, Taxus, Thuja, Tsuga*
4	Temperate BL evergreen tree (TeBE)[1]	*Alangium, Arbutus, Castanopsis, Distylium, Engelhardia spp., Lauraceae spp. (*e.g.*, Neolitsea, Lindera, Persea), Magnolia spp., Olea, Ocotea, Pistacia, Phillyrea, Quercus myrsinaefolia, Quercus Sect. Cyclobalanopsis, Quercus engelmannii, Quercus dumosa, Quercus ilex, Quercus troyana, Reevesia, Symplocos spp., Trigonobalanus*
5	Temperate BL summergreen tree (TeBS)	*Acer, Aesculus, Carpinus, Castanea, Fagus, Fraxinus, Juglans, Liquidambar, Ostrya, Populus, Quercus spp.* (e.g., *robur, pubescens*), *Tilia cordata, Ulmus*
6	Boreal NL evergreen tree (BNE)	*Cupressaceae spp., Juniperus, Juniperus communis, Abies spp., Picea abies, Pinus spp., Pinus sylvestris*
7	Boreal NL summergreen tree (BNS	*Larix spp.*
8	Boreal BL summergreen tree (BIBS)	*Alnus, Alnus glutinosa, Corylus avellana, Populus spp., Tilia spp., Betula spp., Salix spp.*
9	C3 grass (C3G)	*all C3 herbaceous plants*
10	C4 grass (C4G)	*all C4 herbaceous plants*
11	aquatics	*e.g., Alisma, Brasenia, Caldesia, Ceratophyllum, Isoetes, Najas, Nymphaeaceae, Potamogeton, Selaginella, Sparganium, Stratiotes, Trapa, Typha*
12	shrubs	*e.g., Ampelopsis, Asimina, Berchemis, Ceanothus, Corylus Crataegus, Decodon, Eurya, Hamamelis, Ilex aquifolium, Leucothoe, Mahonia, Myrica, Ptelea, Rubus, Staphylea, Styrax, Vaccinium, Viburnum*

[1] This PFT includes both schlerophylous and perhumid temperate broadleaved evergreen trees

Table F2 Model PFTs and corresponding main genera and species represented in the late Miocene fossil record. Shrubs and aquatics were not simulated in the vegetation model.

Figure captions

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

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

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

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

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

Figure C1 a) Biomes (see Appendix B for classification) for the present day control run compared to potential natural vegetation from Hickler et al. (2006), b) dominant PFT in the present day control run, and c) biomes in a) aggregated to megabiomes (see Table C1).

Figure D1. 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, for each PFT and for all PFTs combined.

Figure E1. Agreement Index (AI) values for the 280 ppm and 450 ppm runs for different fractional boundaries of the AI statuses.

Figure E2 a) Histogram of the number of non-absent PFTs (fossil diversity fraction> 0.05) at fossil sites, and b) Histogram of the PFT diversity fractions per PFT per site across all sites, the blue line is from the actual fossil data, the red line is simulated for use in the models to estimate chance agreement, as discussed in the text.

1641

Figure1

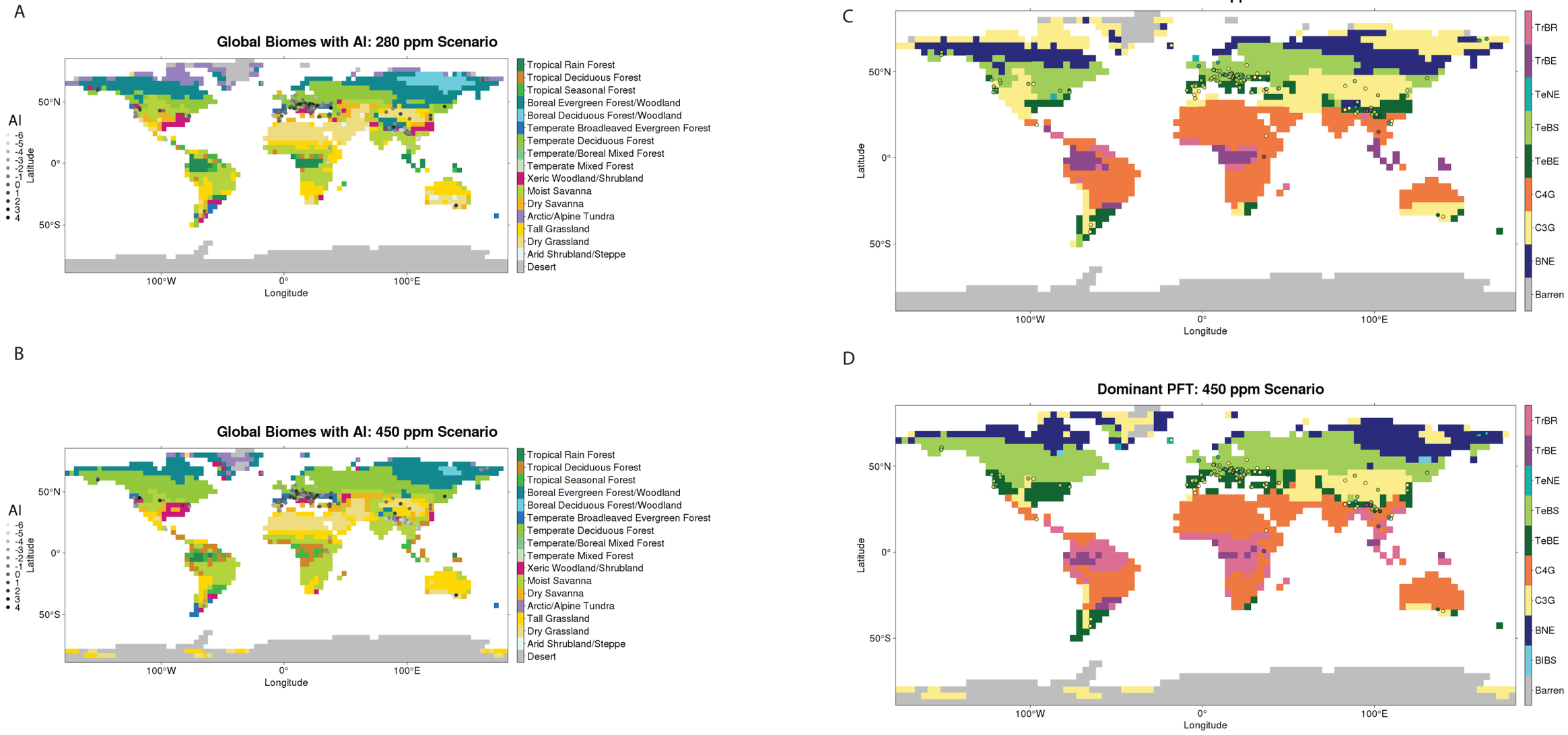

A
Global Biomes with AI: 280 ppm Scenario
AI
-6
-5
-4
-3
-2
-1
0
1
2
3
4
50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
Tropical Rain Forest
Tropical Deciduous Forest
Tropical Seasonal Forest
Boreal Evergreen Forest/Woodland
Boreal Deciduous Forest/Woodland
Temperate Broadleaved Evergreen Forest
Temperate Deciduous Forest
Temperate/Boreal Mixed Forest
Temperate Mixed Forest
Xeric Woodland/Shrubland
Moist Savanna
Dry Savanna
Arctic/Alpine Tundra
Tall Grassland
Dry Grassland
Arid Shrubland/Steppe
Desert
B
Global Biomes with AI: 450 ppm Scenario
AI
-6
-5
-4
-3
-2
-1
0
1
2
3
4
50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
Tropical Rain Forest
Tropical Deciduous Forest
Tropical Seasonal Forest
Boreal Evergreen Forest/Woodland
Boreal Deciduous Forest/Woodland
Temperate Broadleaved Evergreen Forest
Temperate Deciduous Forest
Temperate/Boreal Mixed Forest
Temperate Mixed Forest
Xeric Woodland/Shrubland
Moist Savanna
Dry Savanna
Arctic/Alpine Tundra
Tall Grassland
Dry Grassland
Arid Shrubland/Steppe
Desert
C
Dominant PFT: 280 ppm Scenario
50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
TrBR
TrBE
TeNE
TeBS
TeBE
C4G
C3G
BNE
Barren
D
Dominant PFT: 450 ppm Scenario
50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
TrBR
TrBE
TeNE
TeBS
TeBE
C4G
C3G
BNE
BIBS
Barren

Figure 2

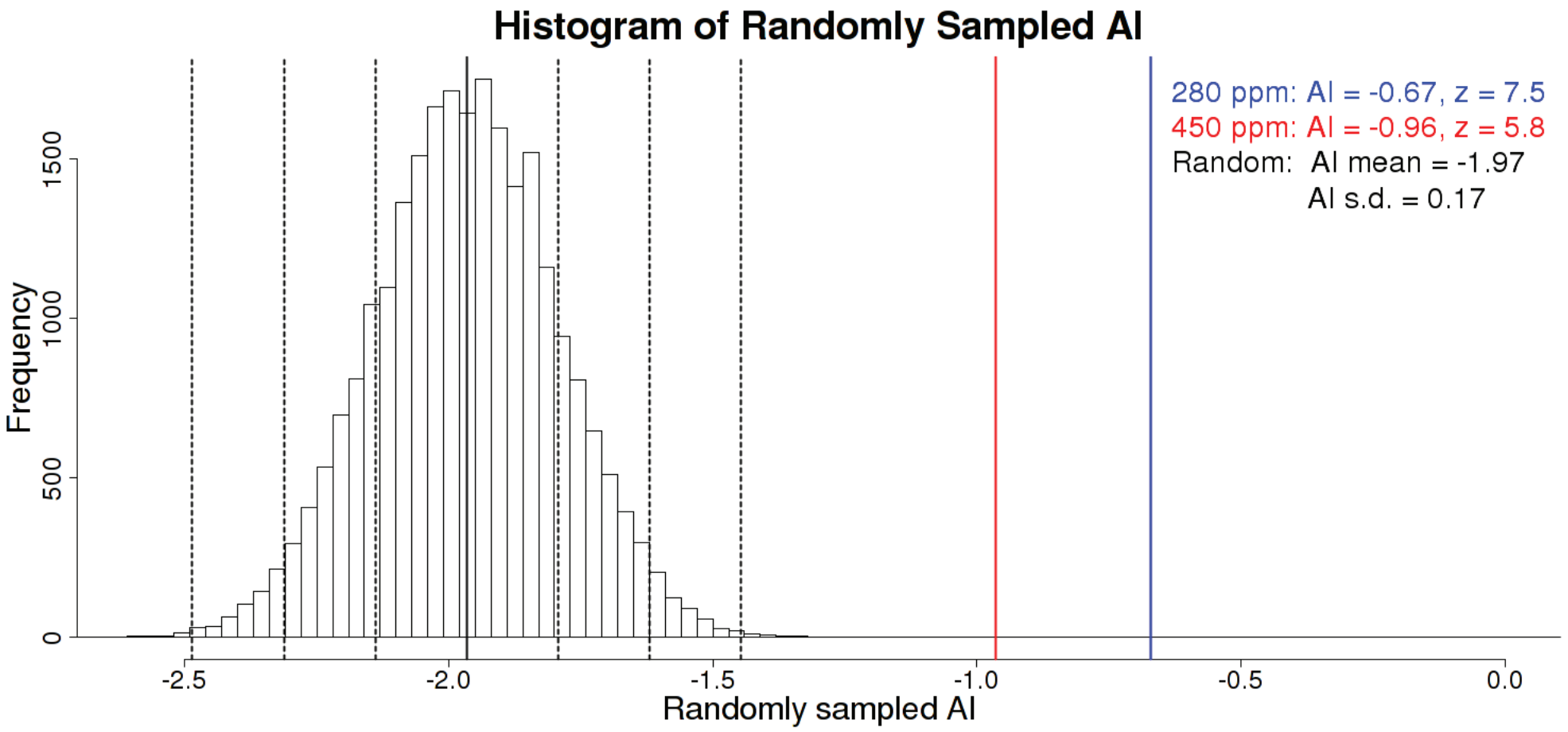

Histogram of Randomly Sampled AI
280 ppm: AI = -0.67, z = 7.5
450 ppm: AI = -0.96, z = 5.8
Random: AI mean = -1.97
AI s.d. = 0.17
Frequency
0
500
1000
1500
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
Randomly sampled AI

Figure 3

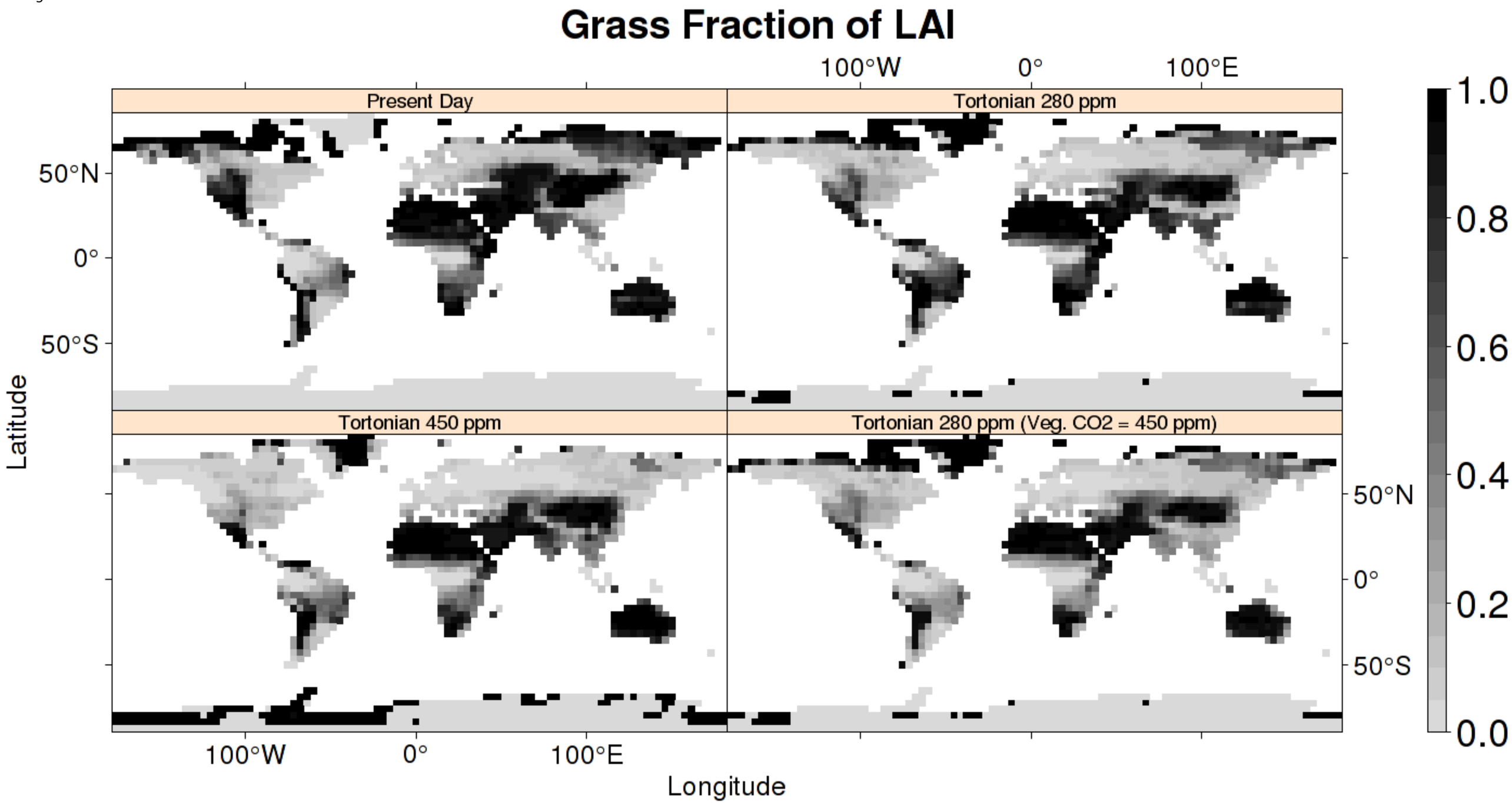

Grass Fraction of LAI
Present Day
Tortonian 280 ppm
Tortonian 450 ppm
Tortonian 280 ppm (Veg. CO2 = 450 ppm)
100°W
0°
100°E
50°N
0°
50°S
Latitude
Longitude
1.0
0.8
0.6
0.4
0.2
0.0

Figure 4

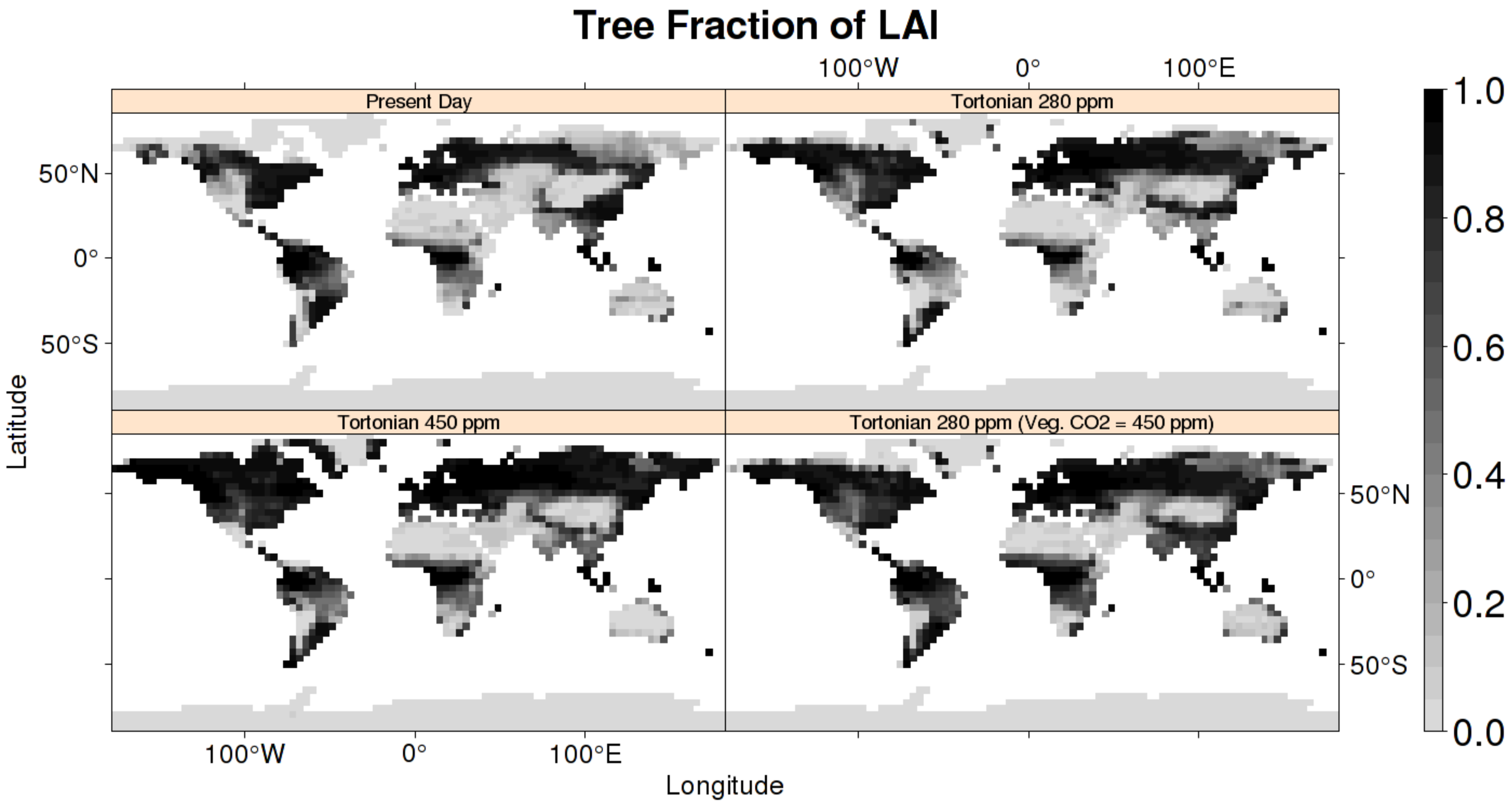

Tree Fraction of LAI
Present Day
Tortonian 280 ppm
Tortonian 450 ppm
Tortonian 280 ppm (Veg. CO2 = 450 ppm)
100°W
0°
100°E
50°N
0°
50°S
Latitude
Longitude
1.0
0.8
0.6
0.4
0.2
0.0

Figure 5

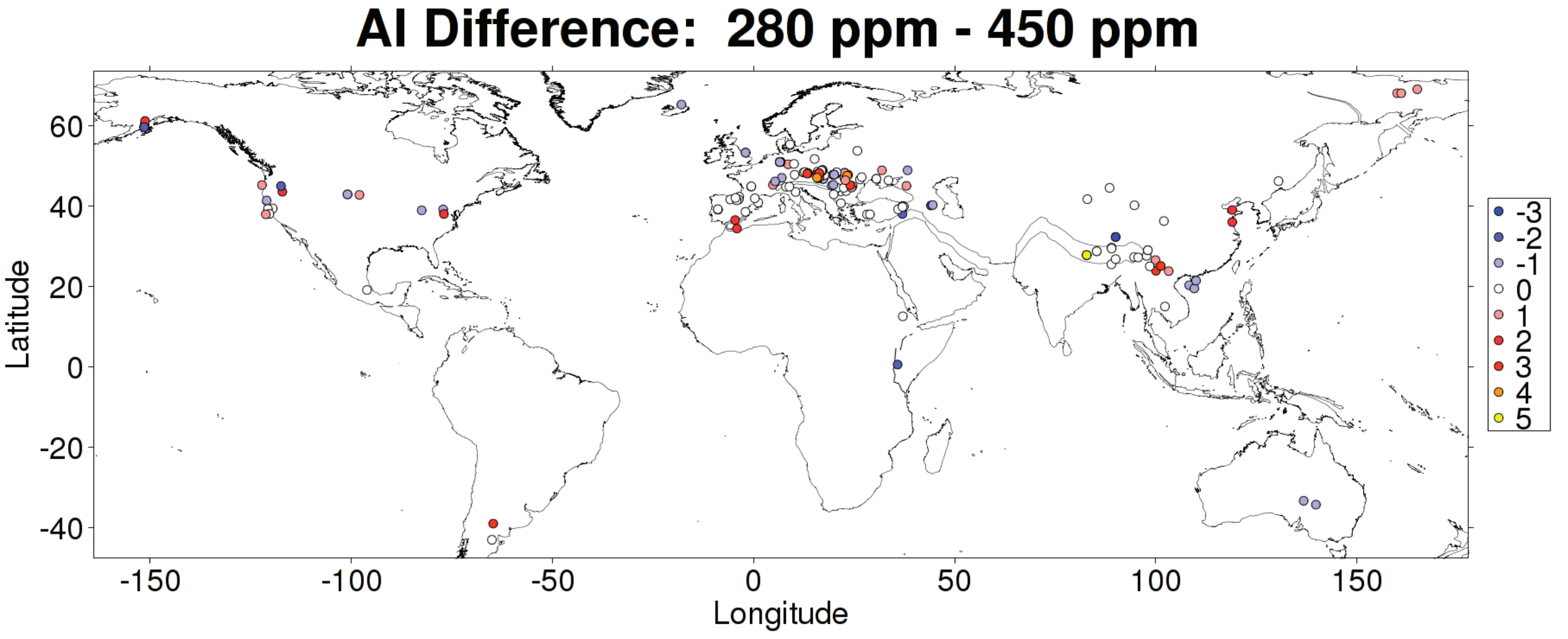

AI Difference: 280 ppm - 450 ppm
Latitude
60
40
20
0
-20
-40
-150
-100
-50
0
50
100
150
Longitude
-3
-2
-1
0
1
2
3
4
5

Figure C1

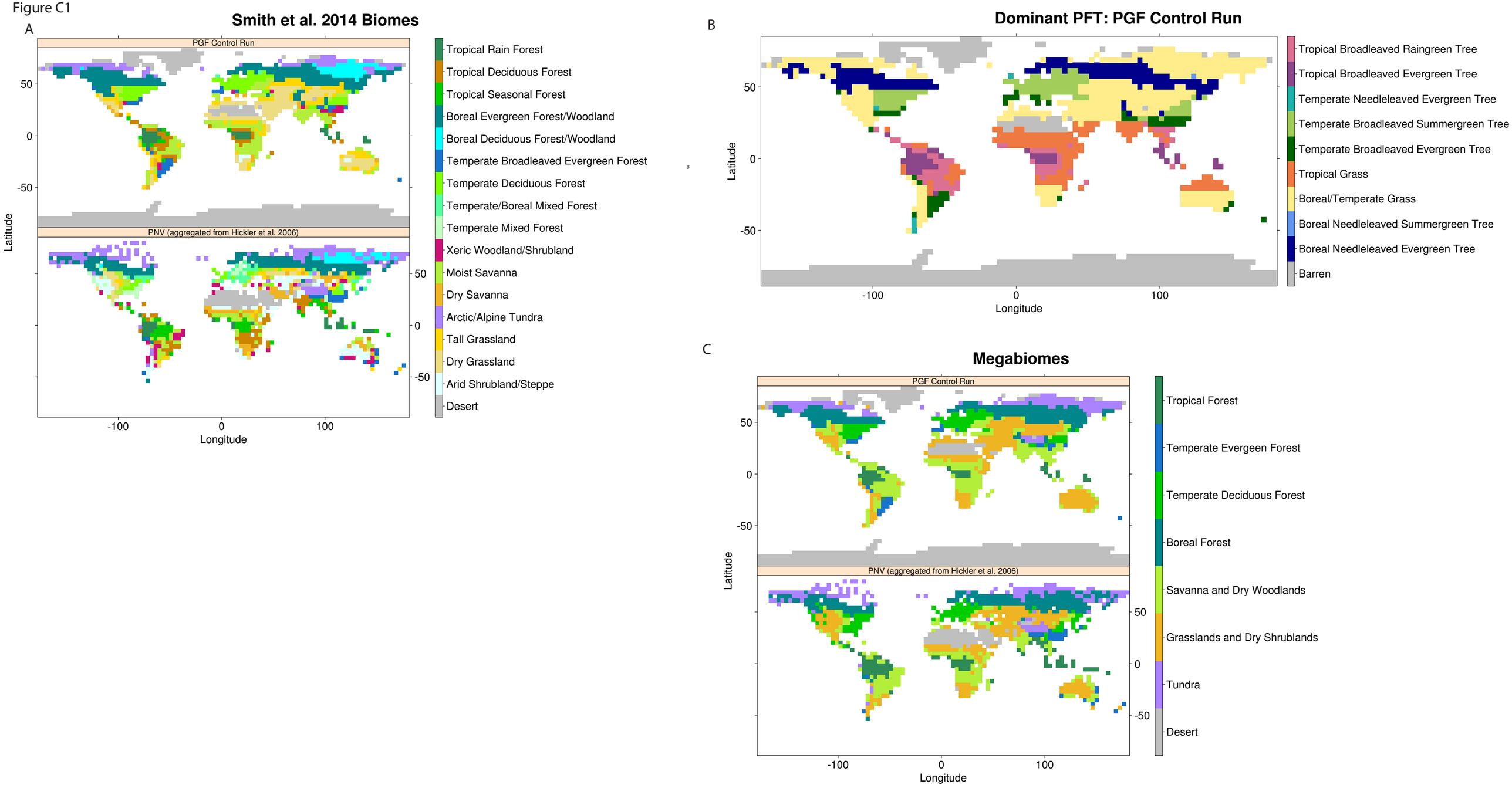

A
Smith et al. 2014 Biomes
PGF Control Run
PNV (aggregated from Hickler et al. 2006)
Latitude
Longitude
50
0
-50
-100
0
100
Tropical Rain Forest
Tropical Deciduous Forest
Tropical Seasonal Forest
Boreal Evergreen Forest/Woodland
Boreal Deciduous Forest/Woodland
Temperate Broadleaved Evergreen Forest
Temperate Deciduous Forest
Temperate/Boreal Mixed Forest
Temperate Mixed Forest
Xeric Woodland/Shrubland
Moist Savanna
Dry Savanna
Arctic/Alpine Tundra
Tall Grassland
Dry Grassland
Arid Shrubland/Steppe
Desert
B
Dominant PFT: PGF Control Run
Latitude
Longitude
Tropical Broadleaved Raingreen Tree
Tropical Broadleaved Evergreen Tree
Temperate Needleleaved Evergreen Tree
Temperate Broadleaved Summergreen Tree
Temperate Broadleaved Evergreen Tree
Tropical Grass
Boreal/Temperate Grass
Boreal Needleleaved Summergreen Tree
Boreal Needleleaved Evergreen Tree
Barren
C
Megabiomes
PGF Control Run
PNV (aggregated from Hickler et al. 2006)
Latitude
Longitude
Tropical Forest
Temperate Evergeen Forest
Temperate Deciduous Forest
Boreal Forest
Savanna and Dry Woodlands
Grasslands and Dry Shrublands
Tundra
Desert

Figure D1

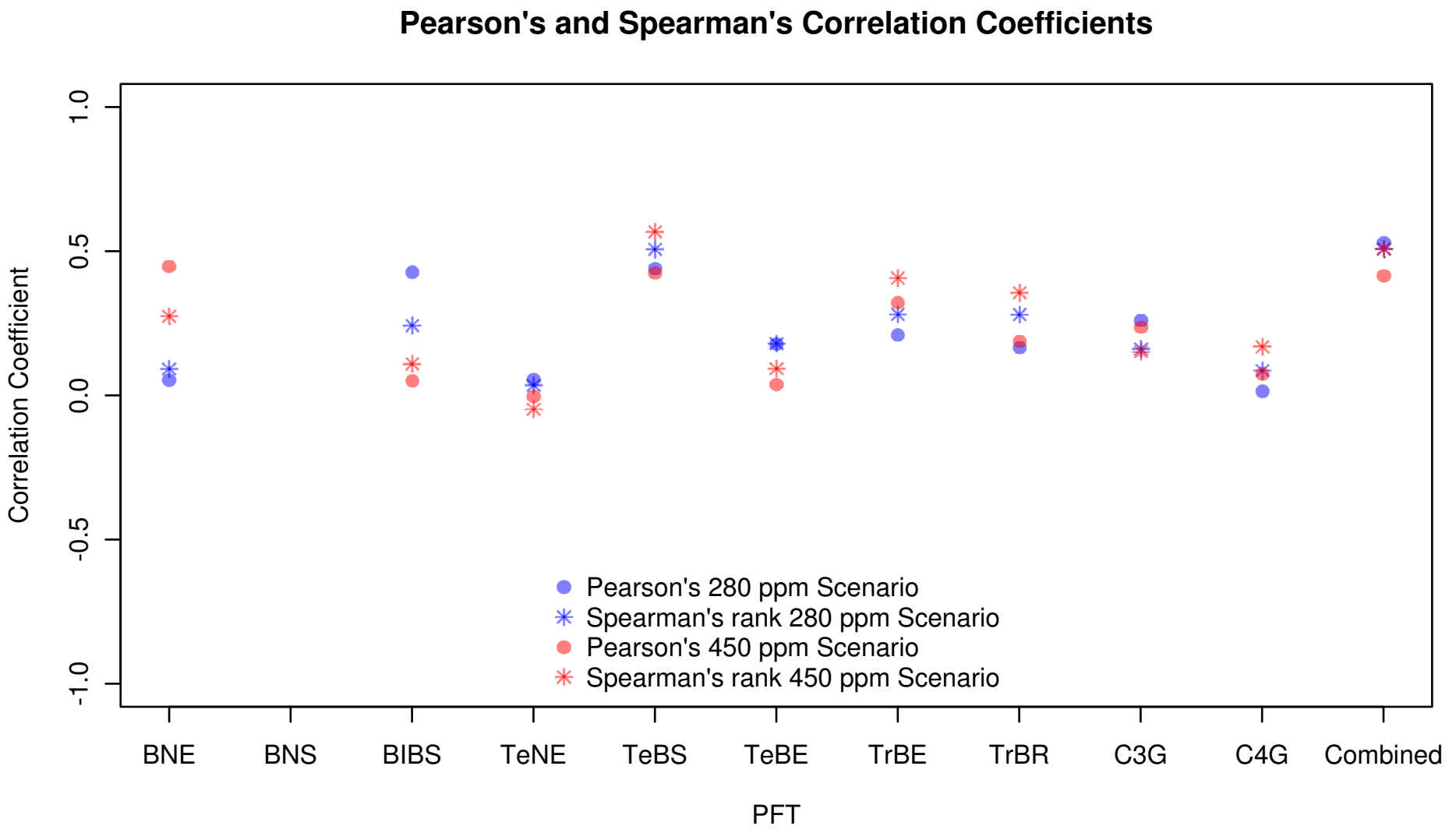

Pearson's and Spearman's Correlation Coefficients
Correlation Coefficient
1.0
0.5
0.0
-0.5
-1.0
Pearson's 280 ppm Scenario
Spearman's rank 280 ppm Scenario
Pearson's 450 ppm Scenario
Spearman's rank 450 ppm Scenario
BNE
BNS
BIBS
TeNE
TeBS
TeBE
TrBE
TrBR
C3G
C4G
Combined
PFT

Figure E1

Agreement Index for a Range of Status Boundaries

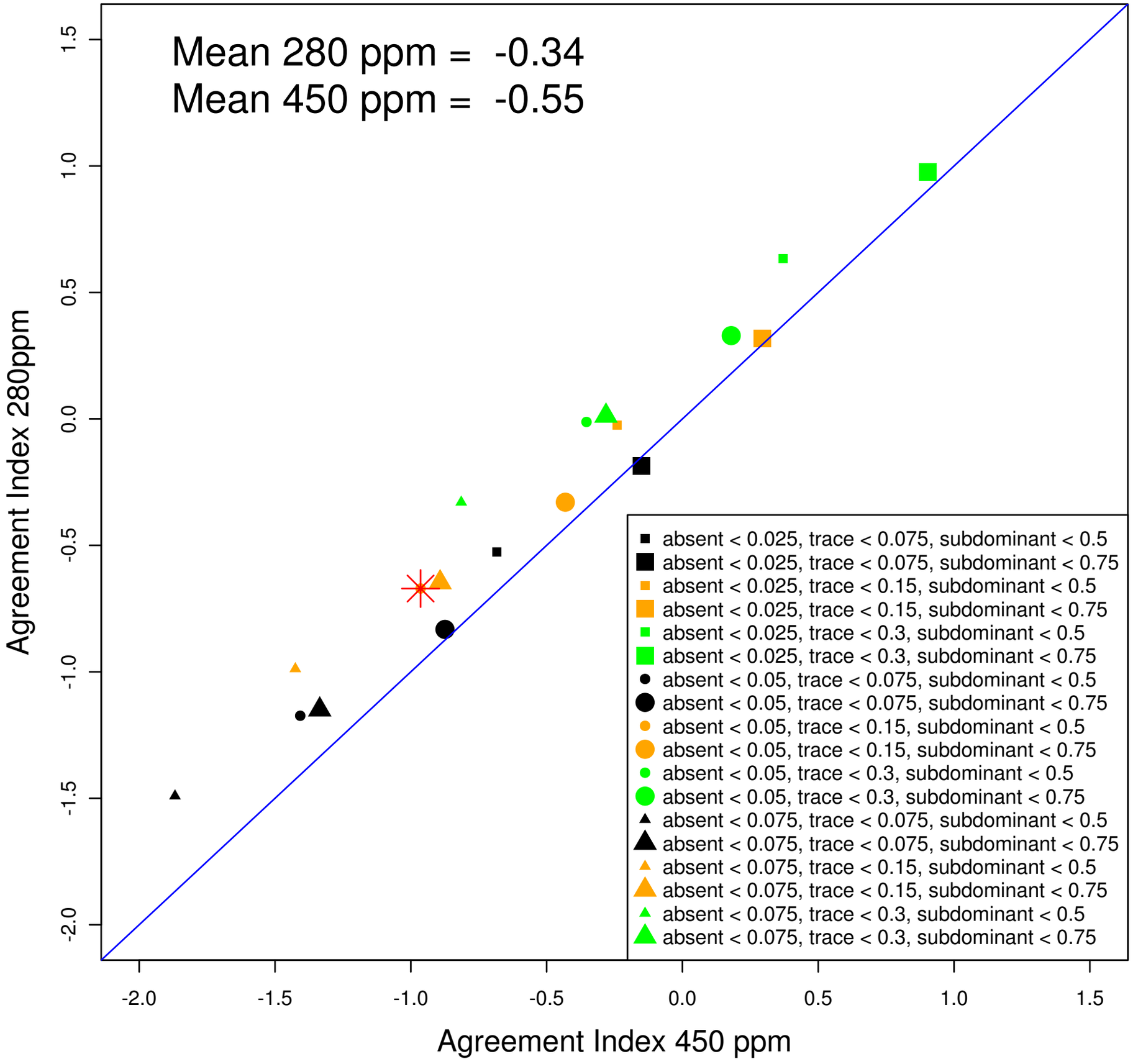
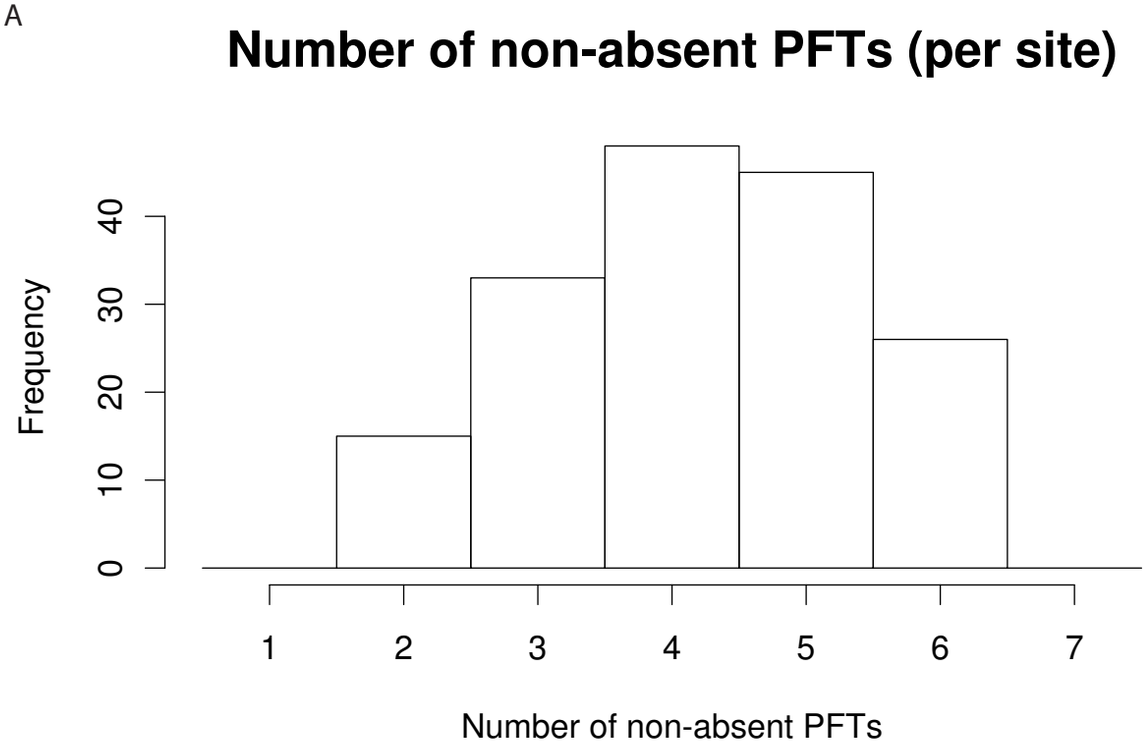
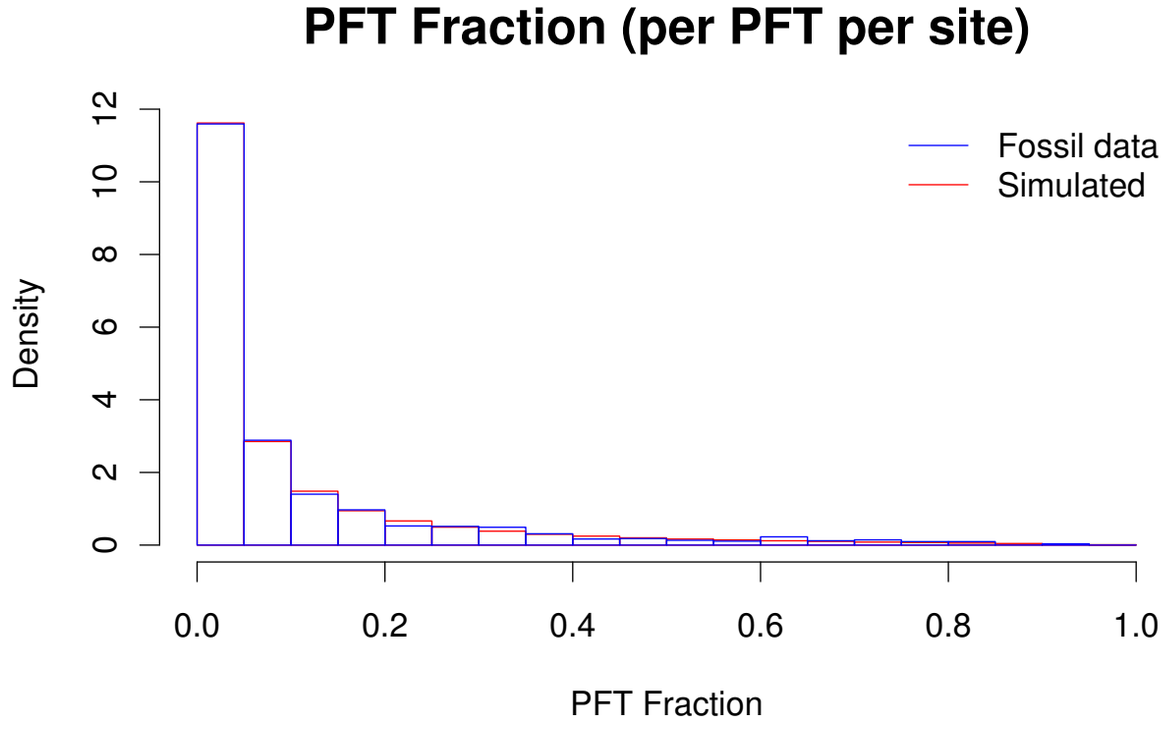


Figure E2

A


Number of non-absent PFTs (per site)
Frequency
0
10
20
30
40
1
2
3
4
5
6
7
Number of non-absent PFTs

B


PFT Fraction (per PFT per site)
Density
0
2
4
6
8
10
12
Fossil data
Simulated
0.0
0.2
0.4
0.6
0.8
1.0
PFT Fraction