Reply: Reviewer 2

General comments
This objective of this paper is to use simulated paleoclimates over the past glacial/interglacial cycle to drive biome simulations, testing those simulations with the pollen data, and then applying the biome simulations to a simple terrestrial carbon/δ13C model in order to explain ocean δ13C variations over time.

The paper used the results of simulations from two climate models, biomes simulated from those climate-model outputs, and the "observed" record of biome variations inferred from fossil-pollen data to check to check the simulated biomes, and finally, calculations of terrestrial carbon-storage variations that in turn govern those δ13C variations.
There is considerable fuzziness in describing which model does what here, and the difference between simulations and reconstructions. The two climate models simulate climate, but not (as they are used here) elements of the carbon cycle (i.e. "interactive vegetation is not included" (in HadCM3), and "interactive vegetation was not used" (in FAMOUS)), so expressions like "The two climate models show good agreement in global and net primary productivity…" don't make sense. Also, the output of BIOME4 is simulated vegetation, as opposed to reconstructed vegetation, which is the product of the pollen synthesis.

We have rephrased throughout the manuscript to try to clearly denote the boundaries of the different reconstructions, simulations and calculations. In particular, the shorthand B4F (BIOME4 simulations forced by FAMOUS climate) and B4H (BIOME4 simulations forced by HadCM3 climate) have been introduced

The general experimental design applied here of using simulated climate to drive biome simulations, testing those simulations with the pollen data, and then applying the biome simulations to a simple model δ13C model emerges slowly in the paper (with the last step not really being discussed until 31 pages into the paper), and so the design might usefully be stated in a "here we use…" fashion in the abstract to get the reader off on the right track.

We have extended the abstract to outline the methods more clearly whilst still trying to be concise. We have also rewritten parts of the introduction to make the design of the paper clearer. At the end of line 159 we added ' In section 2.1 we outline the biomization procedures applied to reconstruct land biosphere changes.' Then at line 163 we added '). Details of the atmosphere ocean general circulation model (AOGCM) simulations are provided in section 2.2.' Line 165 and onwards has been rewritten to ' In section 3 we evaluate biome reconstructions based on these climate model outputs using the BIOME 6000 project (reference), and our new biomized synthesis of terrestrial pollen data records, focusing on the pre-industrial period, 6 ka BP (mid-Holocene), 21 ka BP (LGM), 54 ka BP (a relatively warm interval in the last glacial period), 64 ka BP, (a relatively cool interval in the glacial period), 84 ka BP (the early part of the glacial cycle), and 120 ka BP (the Eemian interglacial).' Finally the last sentence of the introduction has been rewritten to ' Finally in section 4 we apply the biome simulations to estimate net primary productivity and terrestrial carbon storage. Then, using a simple carbon isotope ($\delta^{13}C$) model, we assess the contribution of terrestrial biosphere and carbon storage changes to deep ocean $\delta^{13}C$

over the last 120 kyr and compare this by means of a comparison with deep ocean benthic foraminiferal carbon isotope records, representative for the $\delta^{13}C$ of deep ocean water.'

The data-model comparison (between simulated and "observed" biomes) is relatively lightweight, featuring only a few map comparisons with the relative sparse network of "long" records used here. I was expecting a comparison involving the full BIOME6000 0, 6 and 21 ka data set (which would not be hard to do). It is asserted that the hierarchy of models can successfully simulated the biomes inferred from the pollen data, but this "working hypothesis" as it's called, is never tested, although it could and should be.

This more complete comparison with the full BIOME6000 data for the relevant periods was in fact done, and is largely what is summarised in these subsections, rather than just referring to the much smaller number of longer-record biome sites we synthesise and use elsewhere in the paper for the whole glacial cycle. It was perhaps not clearly signposted that that is what was being discussed, and we have made this clearer in the revised manuscript.

There are three potential sources of data-model mismatches (or indeed accidental matches), 1) the model, 2) the data and 3) the experimental design. For example, disagreements between simulated biomes and observed ("biomized") biomes could be attributable to the hierarchy of models (AOGCMs and BIOME4), the biomization process itself, or to the experimental design (in the application of all of the models). It would be good to discuss those sources and the extent to which each could be influencing the results here.

We have discussed potential mismatches throughout the results and discussion sections. We rewrote section 4 'There is good general agreement between the modelling results and the pollen-synthesis (this paper and BIOME 6000). Below we calculate quantitative changes in the global terrestrial biosphere and carbon cycle, keeping in mind that these calculations carry some uncertainties relating to several mismatches. As is discussed in section 3.1 there are several occasions where the modern biomized pollen data do not agree with actual biome presence; for example Potato Lake and Lake Tulane in North America. In both cases high contributions of Pinus and some other taxa skewed the affinity scores towards drier biomes (grassland and dry woodland). For the past, not knowing whether a pollen distribution is representative for an area, puts restrictions on the biomization method. It is however noted, that in most cases the biomized modern pollen agree well pre-industrial biomes. The models produce some differences in climate and vegetation due to 1) difference in resolution, affecting the biome areal extent and altitude, 2) ice-sheet extent, affecting temperature (section 3.2). If we use the pre-industrial as a test-bed to compare model outputs and pollen (BIOME 6000) reconstructions, there are some biases that can be attributed to biases in BIOME4, the biomization method, and the models limiting geographical resolution.'

Overall, I don't think the case has been made that the hierarchy of models works well enough to use the BIOME4 output for carbon budget calculations. That the approach does work could be demonstrated using the BIOME6000 data, along with multi-

model simulations of 6 and 21 ka; this would also help to evaluate the relative importance of the three sources of mismatch.

See above. We actually do compare the model simulation of the pre-industrial, 6 ka, and LGM with the BIOME 6000 data, and we have stated this more clearly. BIOME 6000 is mentioned throughout the manuscript: lines 147, 166, 614, 647 677, 679 etc. Table 1 lists the BIOME6000 studies that we compared our pre-industrial, 6 ka and LGM model simulations with. However in the text we previously referred to these as high-reolustion biomizations, and to emphasize they are part of BIOME6000 we now explicitly state this in the text.

There is a relatively large author list, which includes some but not all contributors of the original data, and which overlaps a lot with the authors of the individual papers in the Sanchez-Goñi and Harrison (2010) QSR special issue on millennial-scale climate variability and vegetation changes during the last glacial-interglacial cycle. It would be appropriate to provide an indication of author contributions.

Prof. Tzedakis initialized the collecting of long glacial-interglacial records for the Quaternary Quest project with which the first three authors as well as Prof. Harrison and Prentice are associated. The first three authors were responsible for the biomization of the pollen data (Hoogakker) and modelling of FAMOUS (Smith) and HadCM3 (Singarayer, who ran the model with Valdes) and the primary analysis and write-up. Some of the pollen data used in this study came from online databases, some we had access to from the main authors. During the setting up of the biomization matrices Dr. Hoogakker was assisted by the various biomization specialists and palynologists (essentially all the other authors), who critically assessed the procedure and results.

Differences and overlaps with Harrison and Sanchez-Goni (2010):
The Quaternary Science Review Special issue, edited by Harrison and Sanchez-Goni specifically deals with vegetation changes across millennial scale climate change, the Dansgaard-Oeschger cycles. In their paper Harrison and Sanchez-Goni (2010) only discuss certain stadial and interstadial intervals (e.g. their figures 3, 4 and 5). Parts of several of the records that feature in our study are also discussed in papers in Fletcher et al. (2010, Europe), Jiminez-Moreno et al. (2010, North America), Hessler et al. (2010, Africa and South America), and Takahara et al. (2010, for east Asian islands). There are differences in the biomization procedures applied, period covered, but also only a few records are actually shown in those studies. Tropical Asia and Australian records do not feature in this special publication. One major difference too is that all our biomized records are available in the supplementary information. Within section 3.1, where we discuss our biomization results, we also discuss overlaps and differences, with details below:

Line 341 and onwards ' For their study of biome response to millennial climate oscillations between 10 and 80 ka BP Jiminéz-Moreno et al. (2010) applied one scheme for the whole of North America, with a subdivsion for southeastern pine forest. All biomization matrices and scores for individual sites used in our study, generally at 1 kyr resolution, as well as explanatory files can be found in the Supplementary Information.'

Line 359 ' Interestingly, the temperature forest biome has highest affinity scores in a short interval (~15 ka BP) during the deglacation (Fig. 2a). In Jiminéz-Morene et al. (2010) Pinus does not feature in the grassland and dry shrubland biome, but comprises a major component of the southeastern pine forest; hence their biomized Lake Tulane records fluctuates between the 'grassland and dry shrubland' biome and 'southeastern pine forest biome'.'
Line 371 ' Again, In the Jiminéz-Morene et al. (2010) biomizations, Pinus does not feature in the grassland and dry shrubland biome, hence the forest biomes have highest affinity scores in their biomizations.'
Line 382 ' Biomizations for Carp Lake between 10 and 80 ka BP by Jiminéz-Morene et al. (2010) generally look similar to ours, apart from 36, 57-70 and 72-80 ka BP where the temperate forest biome shows highest affinity scores because Pinus undiff. is treated as insignificant in their biomization. Biomizations of Bear Lake between 10 and 80 ka BP are similar to Jiminéz-Morene et al. (2010).'
Line 389 ' Hessler et al. (2010) discuss the effects of millennial climate variability on the vegetation of tropical Latin America and Africa between 23N and 23S.'
Line 415 ' The biomized Colonia record of Hessler et al. (2010) generally shows the same features, apart from an increase in affinity scores for the dryer biomes between 10 and 18 ka BP.'
Line 447 ' Our results are similar to those obtained by Hessler et al. (2010).'
Line 479 ' Fletcher et al. (2010) use one uniform biomization scheme to discuss millennial climate in European vegetation records between 10 and 80 ka BP.'
Line 496 ' Instead of a desert and tundra biome Fletcher et al. (2010) define a xyrophytic steppe and eurythermic conifer biome in their biomizations, giving subtle differences in the biomization records, with the Fletcher et al. (2010) biomized records showing an important contribution of affinity scores to the xerophytic steppe biome. Characteristic species for the xerophytica shrub biome include artemisia, chenopodiaceae and ephedra, which in the Southern Europe biomization scheme of Elenga et al. (2000) feature in the dessert biome and grassland and dry shrubland biome (only ephedra).'
Line 509 ' In the Fletcher scheme characteristic pollen for the eurythermic conifer biome include pinus and juniperus. In our biomization pinus and juniperus contributes to all biomes except for the desert and tundra biome.'
Line 537 ' and Takahara et al. (2010).'

The figures need some work. The key figure is Fig. 2, which shows simulated ad observed biomes, but fuzzes up at the scale necessary to view the results for individual continents.

We have requested for Figures 2a and b to be plotted on separate pages, and also added more details to the Figure caption.

Zooming way in on Fig. 3 suggests that the curves may be “spikier” than they should (i.e. in the data), because it looks like they were constructed with “bevel-joined” line ends (which extrapolate the data, creating the sharp spikes), instead of the more appropriate, but inelegantly named, “butt-joined” line ends.

We have attempted to make the data curves look less spikey, within the limitations of the analysis software used.

Specific comments:
p. 1034/line 5: replace “Global …” distributions” with “Simulated (BIOME4) biome distributions at the global scale” (or something like that).
OK

1034/9: “modelled changes in vegetation” I think this should read “simulated changes in vegetation”—the modelling work got done as BIOME4 was developed; here the model is being applied to generate simulations.
OK

1034/25: “Quasi-periodic” What’s quasi about the periodicity?
We have deleted quasi.

1035/3: “…for the last ~0.8 million years…” What’s special about that interval? Orbital variations have never not influenced climate (and the biosphere, after it developed). It might be better to review the particular variations of climate and its controls over the last glacial cycle than to describe the general ice-sheet, sea-level, CO2, etc. (Quaternary 1010) relationships.
We have rewritten this bit to make it clearer: ' Periodic variations in the Earth’s orbital configuration (axial tilt with a ~41 kyr period, precession with ~19and 23 kyr periods, and eccentricity with ~100 kyr and longer periods) result in small variations in the seasonal and latitudinal distribution of insolation, amplified by feedback mechanisms (Berger, 1978). These are amplified by feedback mechanisms such that for the last ~ 0.8 million years long glacial periods have been punctuated by short interglacials on roughly a 100 kyr cycle.'

1035/9: “productivity and size of the terrestrial biosphere” “Size” could be interpreted a number of different ways, including areal extent, total biomass, etc.
OK, this has been rephrased to 'During glacial–interglacial cycles the productivity of, and carbon storage in, the terrestrial biosphere are influenced by orbitally forced climatic changes and atmospheric CO2 concentrations.'

1035/14: “… the terrestrial biosphere was significantly reduced as forests contracted.” Reduced in what sense? I think the area of the terrestrial biosphere varies rather little over time as icecovered areas seem to be roughly compensated for by exposed shelves. Does this mean instead that forested areas were reduced in area?
We deleted 'as forests contracted'.

1035/15: “21 kaBP” means “21,000 years ago before present”. Just “21 ka”.
Coming from a variety of backgrounds as the authors do, we find the conventions to seem a bit fuzzy as to the precise meaning and implication of various abbreviations. For clarity for all readers, we would like to stress the precise reference point (i.e 1950) of the timescale, so included the BP explicitly, then used ka separately as an SI-type notation for thousands of years. We were unaware that this could be taken with an implicit inclusion of the ‘BP’. Would the reviewer find (kyr BP) a suitable (and explicitly clear) compromise?
.

1036/4: “The data can be viewed through the prism of a global, physically based model that allows the point-wise data to be joined together in a coherent way.” Does that simply mean “interpolation” (which you’re not doing here). Or are you describing how to comparing a sparse network of reconstructions with gridded simulations? In any case this sounds like text from a proposal as opposed to a description of what was done here.
It does not mean interpolation. We have rewritten this sentence to 'The data can be interpreted in the context of a global, physically based model that allows the point-wise data to be seen in a coherent way.'

1036/6: “There are continuous, multi-millenial palaeoenvironmental records… that have not been previously brought together is a global synthesis.” Given the author overlap between this paper and those in the Sanchez-Goñi and Harrison (2010) QSR special issue, this statement is a little surprising. Also, only one kind of palaeoenvironmental data is being synthesized here.
We have rephrases this to ' There are continuous, multi-millennial pollen records that stretch much further back in time than the LGM but they have not previously been brought together in a global synthesis to study changes of the last glacial-interglacial cycle.' Further details of the extent of the overlap with Sanchez-Goñi and Harrison are given above.

1036/14: “We present quantitative estimates of changes in the terrestrial biosphere reconstructed from two atmosphere-ocean general circulation model (AOGCM) simulations over the last glacial cycle.” No you don’t—the “quantitative estimates” come out of BIOME4.
We deleted this and changed the last sentence of that section to ' Finally in section 4 we apply the biome simulations to estimate net primary productivity and terrestrial carbon storage. Then, using a simple $\delta^{13}C$ model, we assess the contribution of terrestrial biosphere and carbon storage changes to deep ocean $\delta^{13}C$ over the last 120 kyr and compare this with deep ocean benthic foraminiferal carbon isotope records, representative for the $\delta^{13}C$ of deep ocean water. '

1036/22: “We assess…” There’s a step missing here. How are biome simulations turned into δ13C values? (Actually the biome-simulation step is missing too.)
We have written a step-wise description of the work carried out, with reference to the various sections where this is being discussed.

1037/3: “Biomization assigns … based on biological and climatological ranges.” To a reader unfamiliar with this process, that might sound like some kind of calibration with climate data is involved.
Deleted ' based on basic biological and climatological ranges.'

1037/16: “megabiome score data…” Why are there blank rows in the spreadsheets? For example, there are pollen data for the Carp. L. sample at 6.12m, but no (mega)biome scores. (Also, why are there two age models for this record?)
We aimed to calculate affinity scores for every 1 ka, with smaller resolution in case the scores across the different biomes were close. We have improved the resolution to 1 ka at Carp Lake. We provide the age models that were originally provided. The

two age models provide some idea of range of ages, illustrating also that there can be large uncertainties.
Added to line 204 ' Sometimes more than one age model accompanies the data, illustrating the range of ages, and also that there can be large uncertainties.'

1038/4: “reconstructions” again
Changed to 'simulations'.

1038/8: “climate averages” “long-term monthly means”?
Rephrase to 'monthly climatologies'

Sections 2.2.2 and 2.2.3: How was land-surface cover specified (or calculated) in the simulations?
It was kept fixed at pre-industrial values, expressed in model variables that are standard for these versions of the MetOffice model (FAMOUS and HadCM3).

1039/10: “biogeochemistry-biogeography model” Should that aspect of the model be mentioned earlier.
Yes, added sentence ' Finally in section 4 we apply the BIOME 4 simulations to estimate net primary productivity and terrestrial carbon storage. Then, using a simple $\delta^{13}C$ model,'.

1039/20: “compare well with NGRIP…”
We have added a couple of sentences with references where comparisons to palaeodata and other models have been made, within this section. The model has been evaluated at high and low latitudes over a variety of time periods.

1039/24: “physically justified ice-sheet extents” Explain.
There is relatively little direct evidence to constrain the extent (as opposed to overall volume) of the northern hemisphere ice-sheets between the Eemian and the LGM, as the proxies on the ground are largely overwritten by the advancing ice, so specifying boundary conditions for this type of modelling work is a significant problem.

For the HadCM3 simulations, the pre-LGM ice sheet areas were obtained by looking at the sea-level change (largely ice-sheet volume) for the timeslice required and taking the ICE-5G extent (a reasonably well constrained reconstruction of the post-LGM ice-sheet collapse) at the time during deglaciation with the same sea-level. This is the extrapolation method we refer to from Eriksson et al (2012). Taken together the HadCM3 timeslices thus show the ice-sheets slowly collapsing in reverse as they grow to their maximum size, which is not physically realistic. The FAMOUS simulations directly used the ice-sheet states for the whole glacial cycle modelled from the icesheet modelling of Zweck and Huybrechts (2005), which produced a physically plausible evolution of the ice. The catch here is instead that such ice-sheet modelling cannot accurately know many of the relevant boundary conditions for the ice, not least the climate they see – it’s a rather chicken and egg problem.

We do not feel that a description at this level of detail is appropriate for the paper – the details are available in the cited papers for those who wish to know – but we have rephrased.

1040/20: “adjusting … to compensate for … biases” and (line 22) “Climate model anomalies …” Is this two separate steps (bias-correction, and then the calculation of anomalies)? What was the base period for the anomalies?
This has been rephrased for clarity. Climate anomalies are produced for a timeslice by subtracting the pre-industrial climate of the relevant model from the climate the model actually simulates for that timeslice. These anomalies are then added to the Leemans and Cramer observations (with the observations interpolated onto the relevant model grid – this appears to be a source of confusion, see later) to produce the climate that BIOME4 actually sees. By using the pre-industrial as a base period for the anomalies and modern observations we neglect differences between preindustrial and modern climate in the models, but these differences are in general small compared to the biases in these relatively low resolution climate models.

1040/22: “temperature and precipitation” What about sunshine, and how was changing insolation handled?
Sunshine anomalies were derived from the models using simulated cloudiness variables. The models did not include variation in either the total output of the sun, nor its spectral composition, which were assumed constant throughout the simulations.

1040/23: “Leemans and Cramer” This implies that BIOME4 was run over the 0.5-degree grid of this data set, but Fig. 3 shows simulated biomes on the grids of the AOGCMs. There’s a big step missing here.
The BIOME4 simulations were in fact conducted on the two different native grids of FAMOUS and HadCM3 respectively, rather than the common higher resolution grid of the climatology data. This has been explained more clearly in the revised manuscript.

1040/27: “model’s” Which one?
Both - corrected

1041/1: “no special correction…” How were modern climate values created for the exposed shelves?
The version of the Leemans and Cramer (1991) climatology included in the BIOME4 distribution includes climate values for these areas. We have not been able to find out their exact provenance.

1041/4: “BIOME4 was forced with appropriate CO2 … (same as used to force the climate model)” Does this mean that Vostok CO2 was used for the HadCM3-driven simulations and EPICA CO2 for the FAMOUS-driven simulations?
Yes, see response to reviewer 1 more details

1042/8: Southeastern? (also in line 15, San Felipe and Potato Lake would commonly be located in the Southwestern US).
Changed to Southeast. Yes.

1042/27: “Recent” as in “present day” or newer than Thompson and Anderson (2000)?
Changed to modern.

1042/29: “… those of the LGM also compare well.” With what?
Move reference of Thompson and Anderson to end of sentence, and replace observations with reconstructions.

Section 3.1: I’m not sure this section serves the paper very well. Each subsection starts with an overview of the location of the sites, but then rapidly becomes anecdotal, describing some aspects of the record for some sites, and different aspects for others. One overall impression I got is that the biomes don’t vary much over time, and another is that there are important differences between the (mega) biomizations here and what was produced in previous studies; neither impression increases confidence about the results.

This is likely more a consequence of the way this section was written than of real issues in the data. As the authors indicate (p. 1041, lines 14-15) only the main results are being presented, but there is no overarching summary—the paper just moves on to the simulation results.

Comparison, where possible, with other studies (generally showing good agreement), has been added, as explained above.

The paper promises a new synthesis, but all it delivers is a few dots on Fig. 2, and some spreadsheets that list the affinity scores, but not the actual reconstructed biomes that the paper is based on.
A new figure 2 has been added showing the affinity scores against time for all the records discussed.

I wonder if this section could be moved to supplemental information, where a more systematic discussion of the individual records could be done, and replaced in the main text with some kind of summary figure. Alternatively, the reader could simply be referred to the Harrison and Sanchez-Goñi summary article in the QSR issue, along with the individual regional articles in that issue.
We have added a new Figure 2, and the results of our biomizations are compared with those featureing in the QSR special issue where possible. (Editor not keen on adding this to supplement).

I’m going to skip commenting on the reset of this section.
1048/18: “where they disagree…” This paragraph starts out talking about the source codes of the climate models, and so it would be easy for the reader to surmise that the disagreement mentioned here is between the climate models and not between the BIOME4 simulations.
This has been rephrased.

1048/21: “coupled to BIOME4” That’s not really happening here.
This has been rephrased.

1049/6: “Because of its lower resolution…” This is certainly true at the resolution of the GCMs, but earlier the experimental design was described as including the “apply-the-anomalies” approach to the 0.5-degree Leemans and Cramer data set

(and repeated on p. 1050/line 6), so presumably the modern “high-resolution” spatial climate variations are also present in the input data for BIOME4.
See above, concerning the grids used for BIOME4. This is what we did.

1049/13: “difference in temperature…” When? At present, or over the course of the climate simulations? (Same question for precipitation…)
This difference is present over much of the simulations - rephrased

1049/21: “warm bias” Again, when?
At the LGM – these has been rephrased.

1049/21: “Millennial-scale cooling events…are not features of our model runs…” Does this mean that they were not simulated, or that the experimental design of the climate simulations did not include the appropriate forcing?
We did not explicitly force the models to simulate millennial scale climate change, and no events of this type were spontaneously produced in the simulations

1049/28: Replace “modelled reconstructions” with “simulated biomes”.
OK

1050/4: The caption for Fig. 2 should point out that it shows both simulated and reconstructed megabiomes. The caption should also explain what we’re seeing on the grids of the two models.
We rephrased this to ' Reconstructed biomes (defined through highest affinity score) superimposed on simulated biomes using FAMOUS (left) and HadCM3 (right) climates for selected marine isotope stages (denoted in ka BP).'

One simulated biome at the grid point? The modal simulated biomes (on the 0.5-degree grid) within the area represented by each model grid cell?
See above, re: grids for BIOME4

1050/29: “additional warmth and sea level” “higher temperature and sea level”?
OK

1051/5: “Differences between our pre-industrial megabiome reconstructions [read “simulated biomes”] only arise from the way the pre-industrial climate forcing [Leemans and Cramer, right?] has been interpolated onto the model grids.” How was the Leemans and Cramer data interpolated onto the model grids? More to the point, why was it necessary to do that? The anomalies of the pre-industrial simulated climate relative to themselves are zero, so there shouldn’t be any difference in simulated biomes, unless something that hasn’t been explained is going on.
See above, re: grids for BIOME4

1051/25: “… on the scale of the climate-model gridboxes.” This makes me think that the biomes are being simulated only for each model grid point, and not for each 0.5-degree grid point in the Leemans and Cramer data set via the “apply-the-anomalies” approach. If that’s the case, the poor agreement throughout between simulated and reconstructed biomes makes sense.
See above, re: grids for BIOME4

1052/9: “… this comparison gives reasonable support to our working hypothesis….” That hypothesis is testable, and indeed should be. It looks to me like there are as many sites with inferred biomes that differ from the simulated biomes as don’t. If you can’t convincingly show that biomes inferred from “modern” pollen data match those simulated by observed climate, then why should we believe the results for other times?
We do show this; through comparison with BIOME 6000. Discrepancies mainly occur near mountainous regions, and are discussed throughout the text.

1052/13: “For both the mid-Holocene and LGM periods, the high-resolution biomizations of the BIOME6000 project (see Table 1) provide a better base…” The same is true for the present.
Yes, as explained earlier we did do this comparison and have added some text.

1052/19: “a greening”?
This has been rephrased.

1052/24: “weak precipitation”?
rephrased

1052/25: “FAMOUS shows a smaller reduction”?
rephrased.

1052/27: “regional biome reconstructions” Do you mean yours here or the BIOME6000 ones?
BIOME6000 – this should now be clearer from the introduction to the section

1052/28: “magnitude of the rainfall” The magnitude of the rainfall or of the rainfall anomaly?
The anomaly

1053/5: “wetter anomalies” Wetter than what? (And it would be better to talk about changes in precipitation as opposed to “wetness”.)
rephrased.

Sections 3.3.2 and 3.3.3: As was the case for the present day, it looks to me that (in the absence) of any quantitative measures, there are as many disagreements and agreements. It is asserted that the more abundant biome data from BIOME6000 shows that “… there is again good general agreement between the two different model reconstructions and the regional biomizations of the BIOME6000 project.” (p. 1053, line 27) but there’s no real evidence that such is the case.
As described above, this comparison has in fact been done and is described in sections 3.3.1 – 3.3.3.

1055/11: “The similar model-based reconstructions…” Similar to what? The LGM simulations? The biome reconstructions?
rephrased.

1055/23: “realistic two-dome pattern” Citation?

Rephrased to describe the ice differences more usefully.

1055/26: “limited vegetation extent” “Vegetation” in this paragraph seems to be equated with tree cover here, but was just used above in the context of land not covered by ice.
rephrased.

1055/27: “wetter climate in HadCM3” Wetter than what?
Wetter than FAMOUS - rephrased

1056/1: “cooler in FAMOUS” Cooler than what?
Cooler than HadCM3 - rephrased

Sections 3.3.3-3.3.5: The main “take away” message I get from these sections is that there is almost no change in simulated or reconstructed biomes over this 40,000 year-long interval, with the ice/land mask accounting for most of any change in the simulations. Is that right?
No, I don’t think that’s really what we’re saying. It’s true that the reconstructed biomes from our pollen records don’t show very much change over this period. Only two of these sites change their highest affinity score between 21 and 54 ka, and only really one between 54 and 64 ka. The global biome simulations however do show significant changes, especially between 21 and 54 ka – see figure 3 (was fig 2 in the first draft). In particular, differences in the different climate anomalies from HadCM3 and FAMOUS through this period can be seen to be significant in their impact on how the simulated biomes evolve. This is true well away from the immediate area of the icesheets, which are specified quite differently for the two models at 54ka.

However, the geographical locations, and sparseness of the pollen records that we have cannot show the quivalent evolution of these biomes over this period, or tell us which of the simulations is more realistic. We have tried to emphasise some of these points.

1056/25: “similar affinity scores to the 64 ka..” “similar affinity scores to those at 64 ka”?
OK .

1056/26: “they are sparse” Sites in general, or those with similar affinity scores at 64 and 84 ka?
Rephrased to 'there are not many sites'.

1057/1: Warmer than what?
The climate at 84 ka BP is simulated as generally warmer than 64 ka BP - rephrased

1057/7: “poorly modelled Mediterranean storm-tracks…” What’s the basis for that assertion?
Mediterranean stormtracks have been shown to be poorly modelled in GCMs of this era/resolution – see e.g.
Brayshaw, D. J., Hoskins, B. and Black, E. (2010) Some physical drivers of changes in the winter storm tracks over the North Atlantic and Mediterranean during the

Holocene. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368 (1931). pp. 5185-5223

1057/9: “Although there are still differences….” I don’t understand “still”
Deleted still.

1057/12: “larger areas of forest”?
Rephrased.

1057/13: “a dry anomaly … that reduces vegetation” Again I think you’re equating “vegetation” with “forest”.
Deleted that reduces vegetation, especially in the HadCM3 reconstruction.

1057/25: “regional climate feedbacks” Explain.
This would take a few sentences to explain and would take away attention of our main message and have therefore rephrased this to ' The affinity scores for temperate forest are almost as high for this site, and neither climate model has the resolution to reproduce the local climate for this altitude well (Bush et al., 2010), although both do reflect dry conditions near the coast here.'

1057/28: “in line with … each other” I’m not sure what this means. Simply “both models”?
Rephrased.

1058/3: Both models increase the extent of their tropical forests…” Does this refer to the BIOME4 simulations? Throughout this section the climate simulations from the GCMs and the biome simulations from BIOME4 keep being conflated.
Rephrased

1058/11: “Quantitative estimates … can thus be drawn…” Yes, but are they meaningful?
Error estimates for this have been calculated and can be seen above in the reply to reviewer 1.
1058/16: “their overall effects” Overall effects of what? From proximity, “their” would refer to “areas and periods with significant regional differences” but that doesn’t make sense.
Changed to 'the effect'.

1058/18: Fig 3. There are three curves shown in each panel. I’m guessing “_S” means shelves and “_NS” means no shelves, but this isn’t explained
The naming conventions throughout have been changed for greater clarity - the new versions (N4F, B4H and B4H_NS) have been explained more clearly

1058/23: “The changes in atmospheric CO2 levels … are common to all BIOME4 runs.” Two things: 1) CO2 changes over time, so it makes no sense that the same levels were used for all runs. 2) CO2 levels presumably don’t vary within a single simulation.
rephrased

1059/3: “FAMOUS also neglects the additional area of land ...” What’s the argument here?
Simply that the FAMOUS-forced BIOME4 run has a smaller area available to be colonised by vegetation than the HadCM3-forced run does at certain periods

1059/5: “global total areas of biomes”?
Added 'of biomes'.

1059/14: “several sites (Fig. 4)” Curves for only one site area plotted in Fig. 4.
A new figure has been made that shows scores for relevant biomes for all sites

1059/16: “~ 70 to 75 PgCyr-1” Which is which? (Later you discuss the NPP values simulated by the two different simulated climates.)
We have now explicitly given the values as estimated separately for B4F and B4H in the text.

1059/23: “… BIOME4 is driven solely by an observational climate dataset…” (for the PI), so on
Added 'for the pre-industrial'.

line 28, the “lower resolution topography” being referred to is that in the 0.5-degree data, right? I find it hard to believe that there is enough smoothing in those data (relative to elevation in the real world) to account for all of the positive NPP bias.
See above, re: grids for BIOME4

1060/6: “In the LGM simulations….”
Rephrased 'The LGM simulations ..'

1060/20: “Further analysis with HadCM3 suggests…” What kind of analysis?
This sentence has been deleted and further detail added towards the end of the paragraph. We were able to separate out the impact of continental shelf exposure, CO2 fertilization, and CO2 forcing of climate by sensitivity experiments with BIOME4 driven with modern or time-slice appropriate CO2, as well as excluding/including the continental shelf areas in global total NPP calculations.

1060/28: “Some differences in the timing of some events… are apparent…” What events?
By events we mean the timing of peaks and troughs on multi-millennial time-scales. These differ somewhat between B4F and B4H. This is mostly a result of the different CO2 forcings used. The text has been changed to reflect this and we have deleted the perhaps misleading term ‘events’.

1061/7: “lower NPP” Than what?
lower NPP compared with pre-industrial times; this has been changes in the text

1061/22: Prentice et al. (1993). Not in references.
Added.

1062/13: Wang et al. (2011). Not in references.

Added

1062/11 – 1063/2: This discussion is really methods, not results. Are the turnover times for different biomes tabulated anywhere? Where does the exponential decay multiplier come from?
Turnover values are now given in Table 3. The decay multiplier is a global, generic estimate from Mahecha et al. 2010, and corresponds to a Q10 of 1.4

Mahecha, Miguel D. and Reichstein, Markus and Carvalhais, Nuno and Lasslop, Gitta and Lange, Holger and Seneviratne, Sonia I. and Vargas, Rodrigo and Ammann, Christof and Arain, M. Altaf and Cescatti, Alessandro and Janssens, Ivan A. and Migliavacca, Mirco and Montagnani, Leonardo and Richardson, Andrew D., Global Convergence in the Temperature Sensitivity of Respiration at Ecosystem Level, Science 2010 doi: 10.1126/science.1189587

1063/3: "The differences in modern NPP by biome between HadCM3 and FAMOUS (related resolution differences…" Please explain. Is "modern" different from "PI"? If so, there's a whole set of simulations that haven't been described anywhere (see also comments about p.1051). If not, why should there be differences?
Pre-industrial was indeed meant. Differences are down to the fact the BIOME4 is run on the two different model grid/resolutions, with some differences in soil properties, atmospheric CO2 levels and how accurately the minimum annual temperature could be calculated from the data available for each model. The resultant global scale NPP calculated is sensitive enough to these differences in each model to make it worth addressing each separately, we feel.

1063/14: "greater retention" Retained from what? (Sounds like from present…)
We intended this to mean that a greater area of forest biomes was maintained going into the last glacial with B4H due to its wetter/warmer climate. The sentence has been changed to clarify this.

1063/23: "greater level of periodicity" I think you're confusing the amplitude with the presence or absence of variations at the ~ 23 kyr time scale.
We have altered the wording.

1063/26: "For the biome scores … (Fig. 3)." Figure 3 shows simulated biome areas.
Rephrased

1063/28: "The largest impact…" On what? The areas? The periodicity of the variations?
We have altered the wording to clarify that we meant the largest contribution to the 23-kyr variations is…

1064/8: "… because other forest types are not compensating periodicities in grassland variation…" No idea what this means.
Deleted this part of the sentence as being confusing and not necessary.

1064/15: I wonder at this point how much of the variation in Fig. 5 is related to the differences in the simulated climates and how much to the turnover times.

We have done some further sensitivity studies on this matter. As noted above, the turnover times derived from the modern carbon/PI NPP for each model are sensitive to the different model setups, and a range of timescales could be equally well justified (see also reply to reviewer1 regarding the NPP-carbon stock equilibrium assumption used in this method). The timescale uncertainty alone feeds through to an uncertainty in terrestrial soil carbon change from the PI to the LGM of order of 10 -20%. The rather large drop in terrestrial carbon reported for the FAMOUS-forced BIOME4 simulations appears to be at the upper end of the possible scale, so the inter-model difference is potentially a little exaggerated in our full results. Even allowing for this, there is still a significant contribution from the different model climates, especially in the smaller scale features in the curves rather than the headline PI to LGM difference. The discussion of figure 5 has been amended in the revised paper.

1064/17: This section has a lot of methods in it, and is rather late in the paper.
There is some method in this section. However as it is a stand on its own feature, derived from the model simulations it features better in its own section; otherwise readers have to make big leaps from this section to a new section with in the methods section 2 (around 25 pages before).

1065/10: “by the model output δ13C for each grid cell” Where do those values come from?
Added 'from BIOME4'.

1065/14: “did not estimate δ13C values” “did not vary (atmospheric) δ13C values”?
We have revisited this section, and are now interpolating atmospheric d13C between the time periods of the available ice core records, so this sentence is no longer relevant.

1065/16: “the calculated δ13C ocean changes would not change” “would not vary”?
Reworded.

1065/20: “total ocean δ13C was calculated for the last 120 kyr (Fig. 6b). Fig. 6b look like it shows anomalies from present day.
The wording has been changed to reflect the fact that we are calculating anomalies.

1066/ 18: “FAMOUS variation is nearly twice the magnitude” Twice the amplitude?
We changed magnitude with amplitude.

1066/21: “deep Pacific δ13C records” Where are those shown?
Rephrased records to stack.

1067/21: “Estimates of global carbon storage reduction are significantly greater if continental shelf exposure is not included…” But the shelves were exposed, so I’m not sure why this is even worth talking about.
Agreed – removed from conclusions

1068/4: “regional climate biases” Biases weren’t ever assessed. The simulated climates differ from one another, but they were never compared with climate reconstructions.

We rephrased this to 'differences in climates between the models…'