Author’s response to reviewer comments:

“Including vegetation dynamics in an atmospheric chemistry-enabled GCM: Linking LPJ-GUESS (v4.0) with EMAC modelling system (v2.53)”

by

Matthew Forrest et al.

We thank the reviewer for giving time to perform a very thorough review the revised version of our manuscript. Here we reproduce the reviewers’ comments in full and address them in turn. As in the previous round, the reviewers’ comments are in black, our responses are in blue. We include proposed alterations to the manuscript to address the reviewers concerns in green.

Review of “Including vegetation dynamics in an atmospheric chemistry-enabled GCM: Linking LPJ-GUESS (v4.0) with EMAC modelling system (v2.53)”

The paper describes the atmosphere-to-land coupling of EMAC to the LPJ-GUESS dynamic vegetation model. Resultant vegetation properties that will become important during the two-way coupling step are benchmarked against observations for offline and two resolutions of the newly coupled LPJ-GUESS in order to determine if discrepancies are due to the underlying vegetation model or climate biases from EMAC. I find it quite refreshing that the authors have focussed on the one-way coupling of part of the land surface as an initial step of ESM development, rather than including the full coupling in one paper. The paper includes a substantial description of future development priorities, which helps frame this analysis in the wider ESM development. It will be interesting to see if this step-by-step approach helps aid these future developments and evaluation of the final ESM. The models' performance seems pretty impressive in the comparisons in this study, which is definitely a good sign for thing to come.

Whilst individual parts of the paper are well explained and easy to understand, the structure of the paper as a whole is quite confusing and could do with some work. I also share some of reviewers 2 concerns around attribution of model discrepancies to the vegetation model, simulated climate or model resolution, as well as the modelling protocol description. However, after going through the manuscript a number of times, I think most of these concerns could also be down to the paper structure. I also have additional questions about the benchmarking methodology which might just need clarification, but could potentially require new analysis. Due to the suggested restructuring and number of general questions, I’m afraid the review is quite long.

We are happy to re-structure the manuscript as suggested and to include clarification of the methodology; we include details below. For convenience we summarise here the main changes in the revised manuscript:

- Restructuring to more standard structure including a “Results and Discussion” section
- Repeated the EMAC T42 and T63 simulations with appropriate N deposition data
- Inclusion of a higher resolution T85 simulation
- Inclusion of GPP comparisons
- Various clarifications and additional information, as requested by the reviewer

One further comment to the reviewer’s use of the term ‘benchmarking methodology’ above. It was not our intention to rigorously ‘benchmark’ the coupled model exactly, but rather to perform some broad-stroke ‘reality checks’ (or perhaps better said ‘initial evaluation’), identify the major biases (and attribute where possible) and investigate the effect spatial resolution. Indeed, we attempted to avoid the term ‘benchmarking’ in the manuscript, although it did creep in a few times which we have now re-phrased. Whilst this distinction is mostly semantic, it does indicate a slightly different intent.

I notice that I seem to be replacing reviewer 2, so I start by briefly addressing responses to reviewer 2 (none of which actually need a fresh response from the authors) before some suggestions regarding paper structure, followed by general methodological comments and ending with specific suggestions to the text.

Responses to reviewer 2.

One of the main concerns raised by reviewer 2 was that the papers focus on natural vegetation at the exclusion of land use coupling. The reviewer included a rather odd comment, suggesting that the authors should consider land use because they are from Europe - almost as if the development of **global** models should be based on our own circumstances or immediate surroundings?! It is perfectly fine to focus on natural dynamic vegetation in this paper, whilst noting (as the authors do) that land use will need including in future developments towards a completed ESM. Dynamics in natural vegetation behaves very differently from agricultural systems, and often require separate consideration (e.g. (Burton et al., 2019)). The paper also accounts for the discrepancy between the simulation of natural vegetation and observations including land use during benchmarking in what seems an appropriate way that is consistent with previous benchmarking studies (Kelley et al., 2013). I, therefore, feel the authors have addressed this point adequately.

We appreciate the reviewer’s comments on this matter.

I do, however, share the reviewer's concern about attribution of simulated discrepancies to LPJ-GUESS model structure, EMAC climate and model resolution; as well as the description of the model protocol, including the prescription of nitrogen, even in the revised manuscript. I have incorporated these into the remainder of my review.

Paper structure

The paper mixes future developments into the introduce and model description and has combined benchmarking methods with results, which at times makes the manuscript hard to follow. On the first read, the paper seems very short on results and discussion, and it's only on subsequent reads that I realised the authors do actually present enough results and make some very interesting discussion points. But these are lost throughout other sections of the paper. It would be much easier to follow the paper's narrative if it was restructured into a more traditional format (i.e an introduction; methods split between model description, modelling protocol, benchmark data and benchmark metric; results and discussion including future work, and then the conclusion). Below, I've highlighted some point that could be moved to more appropriate sections of the manuscript. But I have probably missed some, and I hope the authors can identify some more in the next iteration.

Yes. The previous manuscript structure was admittedly somewhat unconventional as the paper was initially formatted along the lines of technical report, rather than a standard research paper. We have endeavoured to follow the suggestions and re-structure the manuscript appropriately.

Abstract

The parts of LPJ-GUESS that are and are not being used in this study should be explicitly separated in the abstract. Lines 5-8 list a lot of processes that, despite being important in the Earth System, are not being considered for evaluation of this study. This list of processes could be moved into the introduction. Only some processes/couplings in the description of LPJ-GUESS on lines 10-14 are switched on, and not all of those are evaluated.

The abstract is very short on benchmarking methods and results. It should describe what variables are being benchmarked, and give a brief description of attributions of model performance between LPJ-GUESS itself, climate model biases and resolution.

We propose the following revised abstract which we hope addresses the reviewer's concerns.

"Central to the development of Earth System Models (ESMs) has been the coupling of previously separate model types, such as ocean, atmospheric and vegetation models, to address interactive feedbacks between the system components. A modelling framework which combines a detailed representation of these components, including vegetation and other land surface processes, enables the study of land-atmosphere feedbacks under global change.

Here we present the initial steps of coupling LPJ-GUESS, a dynamic global vegetation model, to the atmospheric chemistry enabled atmosphere-ocean general circulation model EMAC. The LPJ-GUESS framework includes a comparatively detailed individual based model of vegetation dynamics including a nitrogen cycle. Although not enabled here, the model framework also includes crop and managed-land scheme, a representation of arctic methane and permafrost, and a choice of fire models; and hence represents many important terrestrial biosphere processes and provides a wide range of prognostic trace gas emissions from vegetation, soil and fire.

We evaluated a one-way, on-line coupled model configuration (with climate variable being passed from EMAC to LPJ-GUESS but no return information flow) by conducting simulations at three spatial resolution (T42, T63 and T85). These were compared to an expert derived map of potential natural vegetation and four global gridded data products: tree cover, biomass, canopy height and gross primary productivity (GPP). We also applied a post-hoc land use correction to account for human land use. The simulations give a good description of the global potential natural vegetation distribution although there are some regional discrepancies. In particular, at the lower spatial resolutions, a combination of low cold and low-radiation biases in the growing season of the EMAC climate at high-latitudes causes an underestimation of vegetation extent.

Quantification of the agreement with the gridded datasets using the normalised mean error (NME) averaged over all datasets shows that increasing spatial resolution from T42 to T63 improved the agreement by 10%, and going from T63 to T85 improved agreement by a further 4%. The highest resolution simulation gave an average NME score of 0.67, just 4% worse agreement than an offline LPJ-GUESS simulation using observed climate data (NME = 0.64). However, it should be noted that the offline LPJ-GUESS simulation used a higher spatial resolution which makes the evaluation more rigorous, and that excluding GPP from the datasets (which was anomalously better in the EMAC simulations) gave 10% worse agreement for the EMAC simulation than the offline simulation. Gross primary productivity was best simulated by the coupled simulations, and canopy height the worst. Based on this first evaluation, we conclude that the coupled model provides a suitable means to simulate dynamic vegetation processes into EMAC."

Introduction

The introduction should mention the importance of attributing biases in simulated vegetation to the vegetation model deficiencies, GCM biases or resolution effects - especially as this probably the most important piece of analysis during benchmarking.

Yes, we agree. We now discuss different sources of biases in the Introduction, see the revised text below.

A lot of the model description should be moved to later sections. For example:

- How climate is aggregated and passed from EMAC to LPJ-GUESS (i.e page 2, line 32 "In both modelling systems…" to end of next paragraph on page 3 line 21) is more relevant to the methods (ie current section 2.3?) and most could be moved there, although it will be worth briefly mentioning that LPJ-GUESS has been used in ESMs before. Also, I'm not entirely sure it's that relevant for this study to describe in detail how LPJ-GUESS is coupled to other GCMs (although maybe the authors disagree on this point…?), and I would suggest making only **brief** comments on past couplings when they use the same technique as the one described in this study.

Yes, this level of detail is excessive and was in part to answer previous reviewer comments. We have significantly shortened the discussion of the other couplings in the introduction and moved (and abbreviated) the details to the Methods section.

- Page 3, line 31 “In addition…” to the end of the introduction could be summarized with the details moved to the discussion/future work.
  Yes, much of this is more appropriate for discussions/future work. We have moved many details to the new Results and Discussions section.

The last two paragraphs of the Introduction now read:

“By bringing together these two modelling systems, our intent is to produce a fully-featured ESM which benefits from the continuous development of both communities. We plan to follow a step-wise model integration roadmap, whereby the coupling between LPJ-GUESS and EMAC is tightened in well-defined, consecutive steps and processes (such as land use) are included or enabled in a consecutive manner. This will allow us to assess the effects of one model on the other, and the effects of the inclusion of new processes, in a step-wise and logical fashion. For our first step, we have chosen to simulate and evaluate the vegetation produced when LPJ-GUESS is forced by EMAC-simulated climate.

When evaluating the vegetation produced in this configuration, there are potentially three sources of error that may contribute to data-model mismatch: poorly constrained parameters values and inadequate representation of the processes in LPJ-GUESS; biases in the climate produced by EMAC (which are expected to have some dependency on the spatial resolution see Roeckner et al. (2006)) and missing processes in LPJ-GUESS (predominantly land use). The issue of missing land use was considered in the design of the evaluation method. To disentangle the mismatches resulting from LPJ-GUESS from those resulting EMAC, we consider a ’stand-alone’ run of LPJ-GUESS in its standard configuration using observed climate data to assess LPJ-GUESS’s implicit biases. To investigate resolution-dependent biases in EMAC, simulations with three spatial resolutions were performed and their performance relative to observed data compared.”

Model descriptions and protocol (section 2 and 3)

Much of the model descriptions include a description of processes/coupling that have either not been implemented or aren’t switched on and therefore not relevant for the results that follow. These are useful to give a wider context to the coupling presented here but should be moved to other parts of the text to help make it much clearer what the authors are actually using in this study. Also, some of the modelling description, coupling implementation and simulation protocol seem misplaced within the methods. For example:

- Unless the authors use fire in their results, all of the paragraph at the bottom of page 5 could be moved to a “future work” section in the discussion. Maybe the authors do use (and during revision, go on to evaluate/discuss) GlobFIRM. In which case the first sentence can be kept.

GlobFIRM is enabled here (as it is a part of the 'standard' LPJ-GUESS set up) so we have retained a reference to it here (but integrated it into the previous paragraph). We now explicitly state in the Simulation Protocol section that fire is enabled in our simulations, where we also note that the output from GlobFIRM is not evaluated as it is a rather simple and (and soon to be out-dated fire model). We discuss alternative fire models in the future work section.

- The last bit of the 1st paragraph on page 6 (line 7 onwards) could be moved to Appendix A, as it has more to do with modification and how to run the model and isn't necessary information to evaluate the quality of the coupling or benchmarking (which the rest of the paper is dedicated too).
  Yes. As these lines were not important at this point in the manuscript, we have moved these lines to Appendix A.

- The first half of the paragraph starting on line 24 on page 6 could be merged with the line 3-5 on the same page to avoid repetition.
  There may be some confusion here due our repeated use of the phrase "coupling strategy". This was clumsy on our part as we should have referred to the "model coupling strategy" (referring to the technical implementation) and the "model integration roadmap" (referring the step-wise plan to integrate the two models). We have amended the text with the new phrasing, and moved the mention on the model integration roadmap to the last section of the introduction (as it motivates the approach taken here), see our response in the Introduction comments above. We also discuss the integration roadmap further in the Further Work section.

- The "next steps" to the end of the following paragraph on page 6, line 26 to page 7, line 7 could be moved to the discussion.
  Done.

- The 1st sentence of the paragraph starting line 16, page 8 should be moved to the model description (section 2.2)
  Done.

Model evaluation.

The second half of the 2nd paragraph on page 9 feels like a discussion on poor model performance... before the results are actually presented! The NME scores actually turn out to be pretty reasonable so this slightly negative statement doesn't just seem misplaced but also takes away from the results that follow. This should, therefore, be moved to the discussion, and phrased in a slightly happier way.

Done. And we are of course pleased to phrase our results in a happier way.

Dataset descriptions and comparisons are combined for each variable. I can see the logic here, but it makes the m/s feel very jumpy, especially given the interactions between e.g. carbon to tree cover, tree cover/height to biome and biomass etc. The authors should consider putting descriptions of comparison datasets and benchmarking metrics into an earlier, more traditional methods section. That way, you can also describe how benchmark datasets are processed in one place (I'm sure I saw "conservative remapping" more than once.)

Yes. We have consolidated the descriptions of the datasets and processing to a new section "Methods/Evaluation data sets".

This should be followed by the comparisons, presented in an order which helps link how model errors in different variables affect one another and are affected by resolution and climate.

Yes, we have moved the description of the results and associated discussion to the new "Results and Discussion" section, consistent with the re-formatting of the manuscript.

With these changes, the former "Model evaluation" section has been entirely deleted and its contents have been split into conventional "Methods" and "Results and Discussion" section as requested.

Discussion and Future work and development plans

There's no discussion section yet, but much of the text identified above could be moved into one. When rewriting, it should be made clear in the text which coupling/model developments are implemented but not yet assessed and what is still to be implemented (this shouldn't require too much effort as the authors have already demonstrated this quite nicely in Figure 1). Once this is done, development and evaluation priorities could be better linked to model deficiencies identified in the results/discussion (as has been done with things like disturbance rates etc).

Yes, this is a very good and logical idea. We have endeavoured to consolidate the details of the results to the new "Results and Discussion" section, and all discussions of further developments and evaluation priorities (particularly in the context of the preceding results) to the new "Results and Discussion/Further work" section.

General comments

Model descriptions

More description of the land surface scheme (outside of LPJ-GUESS) would be helpful. Specifically, as the paper only deals with one-way coupling, what vegetation cover/distributions does EMAC actually see in these simulations? And where are they obtained from? Are there any other, none-dynamic land surface properties that are relevant?

We initially omitted the details because these relate more to the existing aspects of EMAC that we will eventually circumvent in the fully coupled model. However, we agree with the reviewer that these details might be helpful and propose to include the following text:

"As the simulations conducted here utilise only a one-way coupling, EMAC uses its standard land surface scheme which is taken from the ECHAM5 model and is described in detail by Roeckner et al. (2003). Prognostic surface and soil temperatures are calculated with a 5 layer soil model. For the hydrology component, a simple bucket model is assumed, and the water storage capacity is prescribed based on soil type data. A set of land surface data (vegetation ratio, leaf area index, forest ratio, background albedo) has been derived from a global 1 km-resolution dataset for the different horizontal resolutions of the ECHAM5 model (Hagemann, 2002). These data are used to prescribed a climatology of forest fraction (with a constant value) and of vegetation ratio and leaf area index (with a monthly temporal resolution). This prescribed land surface data is used in the model for the calculation of processes such as the interception of precipitation, the snow view in the case of snow-covered surfaces, and for evaporation (bare ground versus vegetated surfaces). Additionally, this data is used in the vertical diffusion scheme and to calculate the grid-mean surface albedo, which depends on a specified background albedo (provided as a constant input data field), a specified snow albedo (function of temperature), the area of the grid cell covered with forest, the snow cover on the ground (function of snow depth and slope of terrain) and the snow cover on the canopy (Roesch et al., 2001)."

For the stochastic processes described, are these processes truly random? Or do they use semistochastic seeded random number generators? I.e if you performed the exact same simulation twice, would you get the same answer?

These processes are semi-stochastic and we use the same seed for every simulation. During development we carefully verified that we get binary identical results from different runs with the same settings (including the case where the model is restarted and when it is not). We add the following text to the LPJ-GUESS model description section to clarify this:

"All stochastic processes are implemented 'semi-stochastically' using a random number generator with a starting seed. This means that for a fixed starting seed, model runs with the identical settings produce identical results."

Modelling protocol

On line 18 of page 6, what is meant by "LPJ-GUESS provides fractional vegetation cover, leaf area index, daily net primary productivity and average height of each PFT to EMAC"? From Figure 1, I'm pretty sure this means that the coupling is technically implemented but not turned on for this study. But the text sounds a bit like EMAC is using information in LPJ-GUESS. The authors should make it clear what is and isn't turned on.

Yes, the reviewer's interpretation is correct and can see the misunderstanding. We propose to simply expand the sentence to read:

“LPJ-GUESS provides fractional vegetation cover, leaf area index, daily net primary productivity and average height of each PFT to EMAC, but these values are not used by EMAC in the simulations presented here.”

We also move the following sentence (“Parameterisations for determining albedo and roughness length are implemented in EMAC, however they are not enabled in the simulations presented here.”) to the “Results and Discussion/Future Work” section where it is more appropriate.

I am a little lost as to what the simulation actually represents? The solar forcing and CO_2 concentration of 367ppm suggest present day (and the authors should state which years this concentration is from). However, nitrogen deposition is from the 1850s - suggesting preindustrial/early historic. What is the reason for the mismatch? I know the authors have said why in the responses, but a better definition of what these runs represent might help explain the mismatch in the paper. And how does Figure C1 show that there is no impact of nitrogen limitation?

The atmospheric CO_2 corresponds to approximately 1999. The mismatch with regard to N deposition data was purely due to miscommunication during the development. We have now repeated the simulations with appropriate N deposition data (from 1990-1999), rendering the figure C1 and discussions of consequences of the N deposition mismatch entirely moot. These have been removed from the revised version of the manuscript.

What time period are the sea surface temperatures from?

The SSTs (and SICs = sea ice coverage) are climatological values from the AMPI2 database, i.e. the mean monthly resolved SST from the years 1995-2000; they include the annual cycle, but do not represent any specific year (neither a special El Nino or La Nina event). This information is now included in the “Methods/Simulation setup” section.

On the whole, the run sounds like an equilibrium run. Was CRU-NCEP detrended to match (both for the spin-up and the final 113-year run)? And overall, what do the runs represent? An equilibrium version of the present day? Or a pragmatic spin up that could be used for further transient runs? Pragmatic is fine - we’re all climate modellers and we know computer resources are too limited to run all the perfect runs we might want. But it would help when interpreting the results to better define the runs.

Yes, these are an equilibrium runs representing approximately the period from late 1990s to the early 2000s, corresponding roughly to the period in which the satellite observations were made. Unfortunately, these runs are unlikely to be utilised in further studies as the saved model state (for both LPJ-GUESS and EMAC) will not be compatible with future model versions.

What resolution was the CRU-NCEP run? If it was different than the T42 and T63 runs, might this have a difference?

The CRUNCEP was run at 0.5 degrees, that default spatial resolution for LPJ-GUESS. It was remiss of us not to include that detail and now include it in the “Methods/Simulation setup” section.

There is no inter-gridcell communication in LPJ-GUESS (ie it can be considered as a ‘site model’ that simulates an arbitrary list of sites) so spatial resolution does not directly affect the processes. The results are therefore only directly sensitive to the input data at an individual gridcell (site) and therefore also to whatever method was used to re-grid/interpolate it.

Was the 500-year spin up for the coupled EMAC-LPJ-GUESS, or was EMAC spun up using a separate protocol before being coupled to LPJ-GUESS? Either is a valid protocol to follow given EMAC doesn't actually see simulated vegetation properties, and it is not entirely evident from the text which is used. Either way, the spin-up protocol for the EMAC part of the model should be described. Did the 100 year period without N limitation follow an initial 500-year spin up? Or does the 100 years with N limitation + 400 years with N limitation constitute the full 500 years spin up?

LPJ-GUESS and EMAC are coupled throughout the spin-up. And the total spin-up is 500 years: 100 without N limitation followed by 400 years with N limitation (i.e. there is no initial 500 year spinup). We have re-written the last paragraph of the “Simulation setup” section to better explain this as follows:

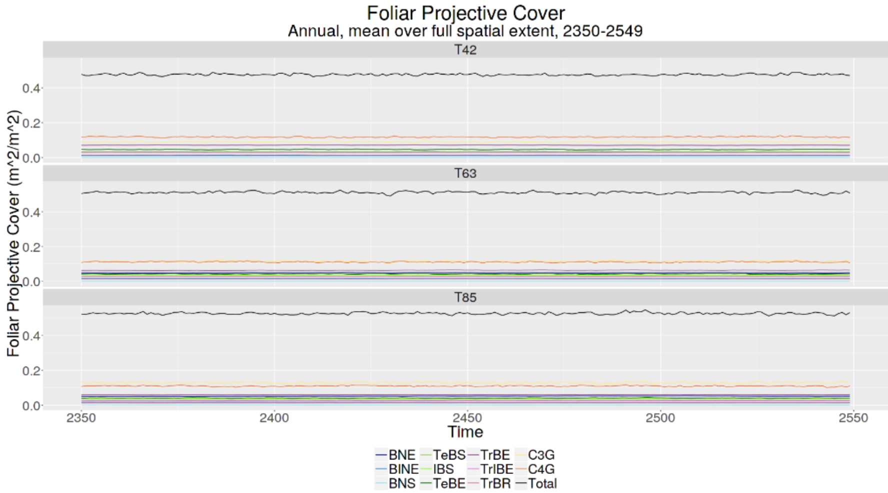
“In all model simulations a 500 years spin-up phase was used to allow the LPJ-GUESS vegetation to reach approximate equilibrium. The coupled simulations used the online EMAC climate during spin-up, and the CRUNCEP simulations the first 30 years (1901-1930) of the CRUNCEP dataset which were detrended and repeated. Simulations followed the standard LPJ-GUESS procedure of starting with ‘bare ground’, ie. no vegetation and no C or N in the soil and litter pools. Having no plant available N present in the soil at the start of the simulation would inhibit and distort vegetation growth if N limitation was enabled. To overcome this, we followed the standard protocol, which is to run LPJ-GUESS for 100 years without N limitation but with normal N deposition to build up the N pools. After 100 years there is sufficient N in the pools, but the vegetation is inconsistent with the desired state as it has been growing without N limitation. Therefore, the vegetation is removed (and the C and N put into the litter pools), and the vegetation is allowed to regrow, this time with N limitation enabled, for a further 400 years. At that time, no significant trends in PFT extension and PFT height were obvious, but the vegetation shows interannual variability as expected.

For the T42, T63 and T85 simulations, an additional 50 years were simulated which were averaged to produce the plots shown here. In the CRUNCEP simulation, a further 113 years (1901-2013) were simulated using full CRUNCEP transient time series. The plots presented here show CRUNCEP output averaged over the years 1981-2010.”

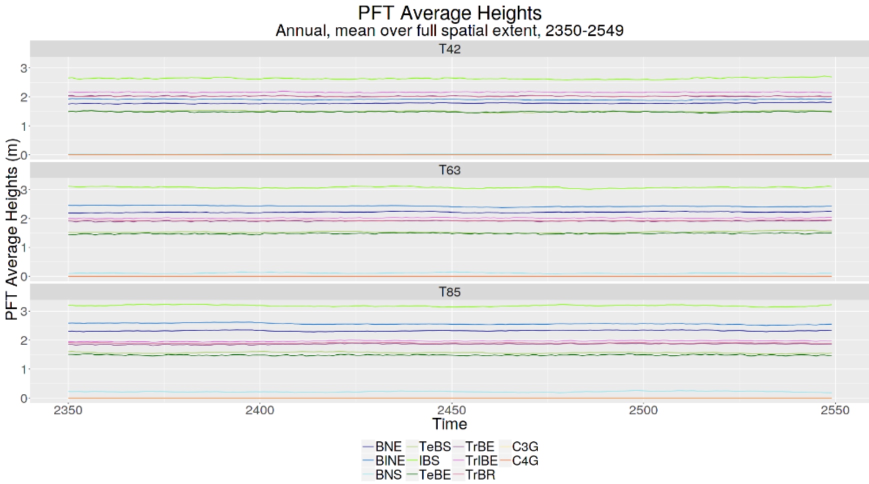
How was the trend in PFT extension and height tested at the end of the spin-up? And does the interannual variability in vegetation refer to extension and height, or other vegetation properties as well? Was the trend in carbon pools tested?

They were examined through simple visual inspection. Since LPJ-GUESS features stochastic processes (establishment of vegetation, mortality and disturbance) and the EMAC climate has internal variability, there will interannual variability in all properties of the vegetation.

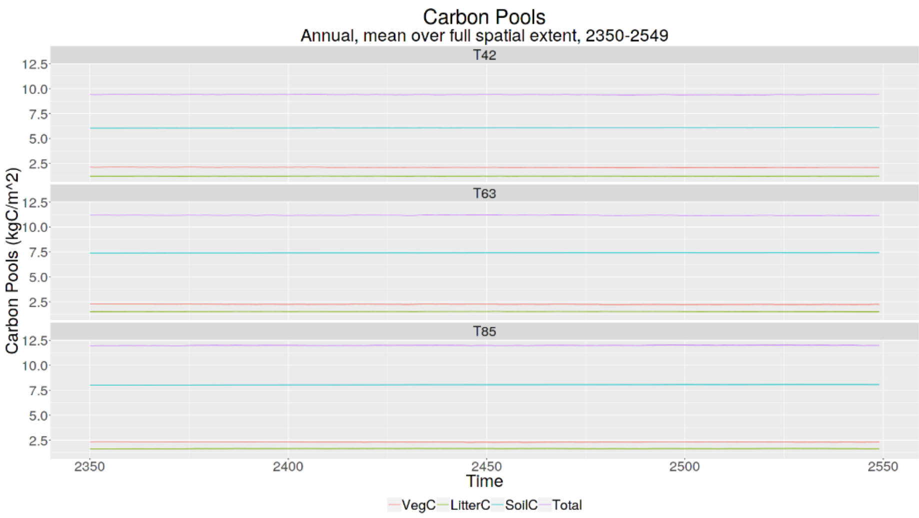
To satisfy the reviewer's curiosity, we include time series plots of the height and coverage of the PFTs, and the C pools for the last 200 years of the simulations below in Response Figure 1.



a)



b)



c)

Response Figure 1. Spatially-averaged times series of a) PFT extent, b) PFT height and c) carbon pools over the last 200 years of the EMAC-coupled simulations.

Null models and metric interpretation

I can't see any reference to null models described in (Kelley et al., 2013; Kloster and Lasslop, 2017). I *think* a potential reason these have not been included is because the benchmarking is used to compare performance across models, rather than quantifying model performance itself. If this is this case, then it should be clearly stated. The 2nd paragraph in section 4, page 9 is probably a good spot to add this. However, I would point out that scores taking into account land use (in Table 2) look pretty good, and using null models may help highlight this.

Yes, we did not include the null models because this paper is not a benchmarking exercise per se, and the benchmarking metrics are indeed used to compare simulations and the effect of the land cover correction, rather than assert a certain degree of performance or improvement over a null model. We propose to add the following text to the new "Results and Discussion/Summary Metrics" section to explain this:

"It should be noted that in this work we are evaluating only the first milestone of model development, and at this point the model is known to be incomplete (particularly with regard to human land use) and no tuning has been performed to either model component. As such, these summary metrics are not meant to demonstrate a particular level of agreement better than some arbitrary threshold, but rather they are included to quantitively evaluate the differences between models runs at difference resolutions and to assess the effect of human land use via a land cover correction factor. They also give a first overview of how the model simulates key features of the vegetated land surface and quantitively indicate (on a normalised scale) which properties are well simulated and which may require additional tuning in future work."

As NME is basically absolute mean error, changes in scores are directly proportional to the distance away from the observations, so can be interpreted as % improvement/degradation in model performance. Maybe this could be explained when introducing NME and used when describing the scores? i.e for tree cover with LUC, T63 represents a degradation in performance of (0.69 - 0.62)/0.69 10.14% compared to CRUNCEP runs.

Yes, this is correct. We now use this handy feature on NME in the "Results/Summary metric" section and explain the calculation in the "Methods/Summary metric" section the following text:

"As NME quantifies the absolute error in the model compared to the data, the relative difference of the values for two models (compared to the same data set) can be considered the relative improvement of one over the other. For example, if one model yields a score of 0.8, and a second yields a score of 0.6, the second can be said to be 25% better than the first, since 0.6 – 0.8/0.8 = -25% ie. a 25% reduction in absolute error. "

Choice of comparisons

Before introducing the datasets, the authors should justify their choices of variables for comparison, particularly why they are important to assess when coupling vegetation to a GCM.

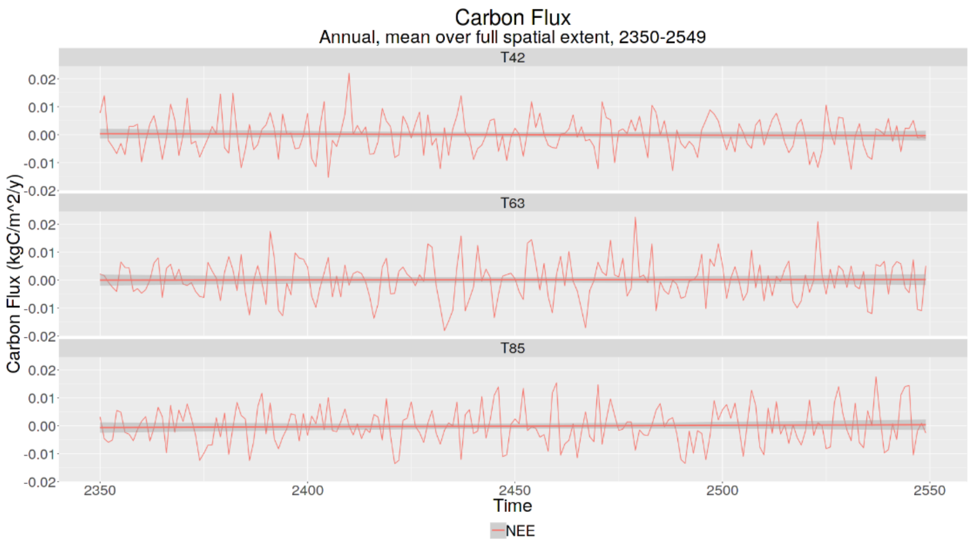
Yes, our reasoning is now included in the new section "Methods/Evaluation data sets".

There is no comparison of important biogeochemical earth system fluxes (i.e NPP/GPP, respiration, ET, methane, aerosols etc). The authors should justify why fluxes aren't considered or else consider adding these comparisons to their analysis. Especially as the title promises coupling to a "chemistry-enabled GCM" where fluxes will be important.

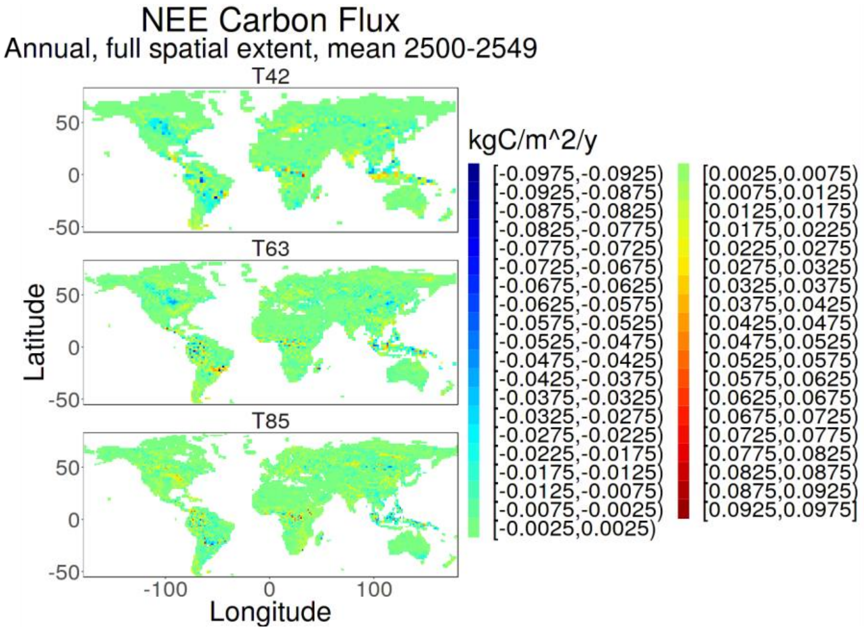
The reviewer makes a good point; we have now included comparison to the Beer et al. gridded GPP dataset. With regards to the other fluxes mentioned by the reviewer, we argue that this first evaluation is not the appropriate point to include such comparisons for several reasons. Firstly, want to focus on the vegetation structure in this work as it is most relevant for the next step of the coupling which will involve bio-physical and hydrological processes (albedo, roughness length and the effect of vegetation fractions and phenological state on water fluxes). Secondly, the land use correction procedure does not translate for biogeochemical fluxes, so their interpretation will be confounded by the missing land use. Finally, the calculation of trace gas emissions (methane, aerosols, BVOCs) was not included in the model setup thus far. We therefore prefer to leave these comparisons for future studies.

As the model should be in equilibrium given the modelling protocol (?) then the authors could also consider demonstrating the model's carbon is in equilibrium as well, i.e net ecosystem exchange is zero - a good basic test for an ESM in equilibrium runs.

Yes, this is also a good point. We have plotted the net ecosystem exchange both temporally and spatially (see Response Figure 2 below). Temporally this does show some interannual variability around zero which is expected (as discussed above) but no trend. Spatially we also see some fluctuations around zero, the 'spotty' patterns reflect stochastic vegetation dynamics processes in LPJ-GUESS whereas the more regional patterns (which are not consistent between simulations) reflect the effects of the internal variability of the EMAC climate.



a)



b)

Response Figure 2. a) Globally averaged times series of NEE and b) spatial patterns of mean NEE for the last 50 years from LPJ-GUESS running within EMAC climate. The straight line in with grey uncertainty in panel a) is a linear fit to the data.

From the description of biome reconstructions in section 4.1, it seems like biome comparisons are being used partially as a way of benchmarking several vegetation properties, including LAI. There are LAI products which could be used for benchmarking if LAI is an important coupling variable (Myneni et al., 2002)?

The biome plots show potential natural vegetation (not present-day land cover), whereas the Myneni et al. data is based on remote sensing of the present-day land surface. Thus, it doesn't make sense to compare the LAI from the version of the model to the Myneni et al. data, two are not comparable. And unfortunately, the land-use correction cannot be meaningfully applied to LAI as crops and pastures still have leaves and therefore LAI. When we have enabled land-use in future iterations of the model, benchmarking LAI will be very relevant.

Vegetation cover comparisons

I know little about biome maps, but (Haxeltine and Prentice, 1996) seems a little bit old. (Oak Ridge National Laboratory, n.d.; Olson et al., 2001) maybe a bit newer? The use of (Haxeltine and Prentice, 1996) over other products should probably be justified somehow.

The Haxeltine and Prentice map is a little bit older, although at 18 years old the Olson map is no spring chicken either! However, potential natural vegetation maps have not advanced so much, particularly as the new wealth of satellite data doesn't help due to the extensive human alteration of land surface. We can see no major methodological advantage of the Olson map over the Haxeltine and Prentice map as both involved compilation of existing maps (most of which, even in the Olson case, are far older than 1995) and a fair amount of 'expert judgement'. Also, as we aggregate to coarser megabiomes the differences between them will be minimised. As the Haxeltine and Prentice map is well known to the much of the DGVM community, particularly those working with LPJ models, and the biome aggregation and classification scheme to derive simulated biome from simulated LAI has already been developed, we prefer to stick with the Haxeltine and Prentice map.

(Kelley et al., 2013) used the Manhattan Metrics (MM) for vegetation cover comparisons. Why was NME used instead? For two item comparisons (ie tree cover vs none-tree cover, as in this paper), I *think* scores obtained using MM and NME would be proportional to one another. If the authors can confirm this is the case (probably with a bit of maths in the response to this review), then using NME will be okay, and could be a better choice as NME is normalised by the variance around the mean of the observations, providing a more intuitive score. But the change in metric should be explained. MODIS MODD44B Collection 6 measures woody cover of a height > 5m. Was LPJ-GUESS tree cover of less than 5m removed? If so, how? Does LPJ-GUESS use shrub PFTS? If so, were they included in tree cover comparisons?

Whilst tree cover can be corrected for deforestation, it is not clear how that should affect the non-tree cover fraction (which would be required to do a MM comparison since the totals of all classes must sum to one). We therefore use tree cover as a single variable so that we can apply the land use correction. In future work, when land use has been enabled, using the MM, including non-woody and bare cover, would be our choice. We propose to include the following text in the "Method/Summary Metric" section to explain this.

“One point where we deviate from their approach is that we use NME for tree cover only, whereas Kelley et al. (2013) use Manhattan Metric (MM) and Square Chord Difference (SCD) to consider the proportions of tree, low-vegetation and bare cover simultaneously. This is because the land use correction applied here represents deforestation by reducing tree cover (see Appendix D), but it does not partition the area that was covered previously by trees into bare or low vegetation cover, which would be necessary for the MM or SCD. We therefore prefer not to benchmark bare or low vegetation cover in this study, and simply apply NME to tree cover.”

Shrubs are not included in the global version of the LPJ-GUESS removed here, and yes, we removed the trees of less than 5m from the tree cover calculation. This was done by computing (in LPJ-GUESS) a separate tree cover output variable excluding trees of less than 5m. We propose to add the following test to clarify this.
“As the MODIS tree cover layer does not include contributions from vegetation under 5m, we added an additional output variable to LPJ-GUESS which sums tree cover from tree individuals taller than 5m only. This variable was used for the comparison to MODIS tree cover.”

Previous cover-based benchmark comparisons also compare herbaceous/total vegetation cover and details on leaf type and phenology (Burton et al., 2019; Kelley et al., 2013; Rabin et al., 2017). Some of these might not all be so relevant for ESM benchmarking, but the authors should either perform them or briefly justify the omission of at least total vegetation cover.
Leaf type and phenology and herbaceous/total cover are explored to a large extent in the biome comparison, which, by considering potential natural vegetation, avoids the confounding effect of the land use. Furthermore, we feel that such a detailed evaluation is not relevant at this stage as we prefer to broader and more robust vegetation properties. To our knowledge this analysis not yet been done even for stand-alone LPJ-GUESS and performing such an analysis would likely require extensive discussion of the performance of stand-alone LPJ-GUESS (including land use), which is not within the scope of this study. However, such work could be considered after land-use has been included in the coupled model.

Fire

The manuscript describes what I *think* is the current fire model (GlobFIRM), as well as plans for future fire model development, at several points in the introduction and model description section. Yet there is no mention of fire in the results or discussion, except a brief comment about possible underestimation of fires effects on canopy height. If fire is important for any of the variables tested (e.g for veg distribution, as it is for (Bond et al., 2005; Burton et al., 2019; Hantson et al., 2016)) or for further development of ESM coupling, then surely it should be benchmarked as well? Simple burnt area and emission benchmarking is described in (Kelley et al., 2013; Rabin et al., 2017) and could be applied.
Yes, fire from GlobFIRM was enabled in this study is it is part of the standard LPJ-GUESS configuration. However, GlobFIRM is rather simplistic and out-dated, and will soon be replaced in LPJ-GUESS by the SIMFIRE-

BLAZE fire model. Thus burnt area and emission benchmarking would not be particularly enlightening and would prefer to address this in a future version with SIMFIRE-BLAZE.

Impact of spatial resolution

The authors suggest at several points that degradation of performance of vegetation properties between T63 and T42 runs demonstrate poorer performance in EMAC at coarser resolutions. It's well established that changes in GCM resolution has an impact on model performance (e.g. (Kuhlbrodt et al., 2018)). However, I'm not sure the comparisons presented in this study lend any evidence to support this. It could be LPJ-GUESS performs worse at coarse resolution, and/or is sensitive to the aggregation of inputted climate information. Unless the authors can justify this statement another way (i.e driving LPJ-GUESS with CRU-NCEP gridded to the two scales of grid - which I'm sure would be much more effort than it's worth? Or making more use of the climate maps in the appendix), I would rephrase this argument to suggest the model as a whole (i.e EMAC+LPJ-GUESS) performs better at the increased spatial resolution, particularly in the first paragraph on page 17 and lines 4 and 5 on page 18.

As mentioned above, the processes in LPJ-GUESS are independent of spatial resolution. From the reviewer's comments it is apparent that this was not adequately explained in the manuscript. We propose to add the following text to the "Introduction" section to clarify this,

"As LPJ-GUESS has no inter-gridcell interactions and no processes are gridcell size/spacing dependent, it has no sensitivity to the spatial resolution at which it is run. Thus, any improvement in the vegetation produced by the EMAC-coupled simulations at higher resolution can only be due to changes in the EMAC produced climate (i.e. reduced biases)."

And also to the new Results and Discussion section where the improved NME scores are discussed:

"As LPJ-GUESS processes are completely independent of spatial resolution, these improvements must be coming from the EMAC-simulated climate."

The reviewer is correct that in principle if climate data is aggregated before being used in LPJ-GUESS, LPJ-GUESS is sensitive to that aggregation and the method used could have an effect on the results. However, that is not relevant here as the climate fields are calculated at these resolutions natively by EMAC, so no aggregation has been done.

Given the above points, we believe that we do provide evidence that coarser spatial resolution of EMAC adversely effects the simulated vegetation, since any worse performance as coarser spatial resolutions can *only* be due to the climate fields provided by EMAC (admittedly the regridding of the evaluation data might have a small effect, but as these were re-gridded from fine resolution to the simulation resolution by the same method, this is unlikely to have a large effect). The most obvious results are the lack of vegetation at high latitudes at low resolution and the consistent pattern of improving NME scores going from T42 to T63 to T85 resolutions.

To explore resolution effects further, I do actually think it would be interesting to see T42 in the other maps. There are clearly some big differences in biome cover in some parts of the world between model resolutions, including some interesting changes in the biomes in carbon-rich Amazon, Indonesian and South Asian forests (changes which I don't think are mentioned in the text?). It would be useful to see where these differences are coming from (including a T42 tree cover, height, and biomass maps and linking these to Figure B1-B3 climate where T42 is already plotted out).

Yes, we are happy to add the T42 plots for tree cover, height and biomass, and to discuss the results in the new "Results and Discussion" section.

Figures

Why is the background in different maps in figures 3-5 blue? Shouldn't they be white to match the part (a) figures?

As we chose white as a neutral colour for 'no difference' in the difference plot, and found that a coloured background (rather than white) conveyed the results more effectively. However, on second thoughts a blue-grey background is not ideal (as the colour scale contains blue), so have switched to a neutral light-grey. We hope the reviewer finds this reasonable!

There are streaks of missing data running across the Sahara and Arabian Peninsula and southern Australia what I think is in the (Avitabile et al., 2016) region of the biomass observations map. What's causing this?

This is data missing from the Avitabile dataset. There was a note about this in the figure caption, and we have added one to the new "Methods/Evaluation data sets" section.

Specific comments

Given the timing of the submission, I wonder if this model will be contributing to CMIP6. If so, the authors may wish to point this out somewhere?

No, we were not working to the CMIP6 timetable, that seemed rather ambitious. Perhaps CMIP7.

Remove most of the instances of "state of the art". The authors either clearly demonstrate that the models they are coupling are the latest version, in which case the phrase is redundant, or use it when there is no extra justification as to why the model is "state of the art", in which case we are left wondering why it warrants the description.

Done (although we left it in the introduction to describe ESM a category).

Page 2, line 17 is the phosphorus cycle considered in EMAC or LPJ-GUESS? If so, add to the model description.

Phosphorous is not included in either.

Page 2, line 26 add "being" between "actively" and "developed"

Done.

Page 4, line 23/24: “comparatively detailed”. Compared to what?

Compared to models with no cohort or age structure in their vegetation. However, since the subsequent text discusses this point we now simply omit the comparatively.

Page 4, line 29: replace “phenology” with “phenological response”

Done.

Page 4, line 29: C4 is just for grasses right? If so, state.

Yes, and now in the text.

Page 4, line 30: Are any of the woody PFTs shrubs? If so, say so. If not, don’t worry.

Unfortunately, there are no shrubs as of yet in the standard global LPJ-GUESS configuration.

Page 5, paragraph starting line 5: remove the information of version 3 of LPJ-GUESS. Unless version 3 is going to be used somehow (?), it is not relevant to this study. Apart from that, this is actually a very clear and concise overview of how LPJ-GUESS works!

The references to v3 were superfluous and have been removed.

The overview was in fact taken from a freely available generic LPJ-GUESS model description on the LPJ-GUES website, so we can’t take credit there. We have added a short note to explain this in case of being flagged by automatic plagiarism detection software (even though the text is explicitly an open and non-copyrighted resource).

“The following two paragraphs are modified from a standard model description template which is freely available and copyright free.” (with foot note to: http://web.nateko.lu.se/lpj-guess/resources.html)

Page 5, line 20-22 “However, monthly climate data… distinct rain events”. Is this climate interpolation actually used in coupling to EMAC? Or is it used when driving with the CRU-NCEP in this study? If not, remove.

Yes, it applies only to the monthly CRUCNEP. We now make a note to this effect in the “Methods/Simulation setup” section.

Page 7, line 6-7: remove “Whilst it's not within the scope … dependence on spatial resolution” and if necessary, merge the remainder of the sentence with the next. The paper does later test biases/resolution that affects dynamic vegetation. And it’s self-evident that your aim is to test the coupling and veg dynamics, not the wider GCM.

Done.

Page 8 line 1: state what resolution or T42 and T63 are. (ie no. lat x lon. Deg at the equator or something like that). I know spectral resolutions are a little confusing in these regards, but it would be useful to give an idea of how much coarser T42 is relative to T63. And maybe discuss somewhere what applications each resolution is likely to be used for.

Good point, we include this information now. And we include the following text in the "Simulation set up" section:

"Whilst finer spatial resolution (such as T63 or higher) exhibit lower biases (Roeckner et al. 2006) and so are generally preferred wherever possible; coarser resolutions (such as T42) may be used in situations where the large computational cost prohibits finer resolutions, such as long runs transient simulations of paleoclimate, factorial or sensitivity studies or simulations including detailed atmospheric chemistry calculations."

Page 8, line 8 remove "at no point were external climate datasets used". I think the authors mean that no external climate datasets were used in the T42 and T63 runs. But that's pretty self-evident, so stating it here sounds like at no point in this analysis was climate data used, right before describing how climate data was used.

This was to answer a point from a previous reviewer, who obviously did not find it self-evident. We therefore prefer to keep the clarification. However, we now clarify that:

"Throughout both the **coupled T42, T63 and T85** simulations,"

which we hope alleviates the reviewer's concern.

Page 8, lines 19-22: Explain how Fig C1 shows that PI N deposition does not result in additional N limitation. And additional when compared to what?

We have repeated the simulations with appropriate (1990-2000) N deposition data, so have now removed that text and Fig C1.

Page 9, line 4: This paragraph starts off with a very long list of previous benchmarking of LPJGUESS. It should be reduced to just state that LPJ-GUESS has a long history of development and evaluation, and then pick a couple of key references. And recent ones preferably using the version of LPJ-GUESS used here (if I remember correctly, (Gerten et al., 2004) for example, was pre-GUESS?)

Done, the first paragraph of the " Model Evaluation" section now reads:

"Stand-alone LPJ-GUESS has a long history of development and has been evaluated in detail in previous work (some recent examples include modelled potential natural vegetation and forest stand structure and development (Smith et al., 2014), global net ecosystem exchange (NEE) variability (Ahlström et al., 2015) and the effect of CO_2 fertilisation (e.g. Medlyn et al., 2015)). Here we performed an initial evaluation focused on vegetation state variables relevant to the biophysical coupling between the land and atmosphere in the coupled model setup in order to investigate how LPJ-GUESS responds when EMAC climate is used as the forcing data and to investigate any biases in the vegetation produced.)"

Page 9, line 10-11: Remove “it is beyond the scope…. dynamic vegetation model”. I’m not sure it is beyond the scope to suggest changes to the dynamic vegetation model that might affect/be affected by the coupling. Especially because the authors do later describe some changes (i.e to disturbance rates, fire modules etc) that will change the dynamic vegetation model.

That text has been removed (see revised paragraph in answer to the previous point).

Page 9, line 31 replace “unity” with “1”

Done.

Page 10, line 6: The sentence “Whilst we can’t expect … forced using EMAC climate” seems to be making the same point as the start of the paragraph, and should be moved/merged into somewhere around lines 3 and 4.

Done.

Page 10, line 15: The paragraph starts off a little negative. Maybe just remove everything before the first comma and start the sentence by saying “Knowledge of EMAC biases…”.

Done.

Page 10, line 32: Add “Fig” before “2”

Done.

Page 11, line 24: Briefly explain conservative remapping.

Done, text is included in the new “Methods/Evaluation data sets” section.

Page 11, line 25: removed “as would be expected by a state-of-the-art DGVM”. I’m sure I would expect some DGVMs to do a rubbish job. The fact that you have a combined model which does kind of okay is pretty impressive and shouldn’t be understated.

Done.

Page 12 Please add dataset reference to figure 2 caption.

Done.

Page 15, line 22: remove “(lower is better)”. It’s already been described and the reader is reminded in the table caption.

Done.

Page 15, line 26, 27, sentence stating “For biomass…” states that biomass performance gets worse when accounting for LUC. I might be reading this wrong, but from Table 1, it seems the biomass results are actually getting better with the LUC modification, not worse…?

Apologies, there was an error in the table (now corrected), so the results **do** get worse with the land use correction.

Page 15, line 31, sentence starting "In particular". The authors state that the background disturbance rate could be changed to increase vegetation carbon in the T42 run. However, as far as I can tell, this disturbance rate is also used in the CRU-NCEP run as well (?), where biomass is generally too high.
Apologies, this was also unclear from us, we were not specifically referring to the T42 simulation in this context. We have moved much of this text to the new "Results and Discussion" section and attempted to clarify this. We also included percentage improvement (as suggested above) and the T85 results so the text now and reads:

"The NME scores for all simulations and all evaluation variables are presented in Table 1. In the case of tree cover and biomass, the results are also presented with a land use correction (LUC) factor (see Appendix D). Applying the LUC has a marked improvement on the tree cover NME scores (in terms of percentage reduction of error: 16% for *T42*, 22% for *T63*, 33% for *T85* and 43% for *CRUNCEP*), implying that much of the discrepancy seen between simulation and observation apparent in Fig 3. is indeed due to human land use as expected. For biomass the results are not so clear cut. Whilst applying the land use correction improves biomass agreement in the *CRUNCEP* simulation (by 26%) and the *T85* simulation (by only 3%), it leaves the agreement in *T63* simulations essentially unchanged (0% change) and actually it worsens agreement in the *T42* simulation by 4%.

For the coupled simulations, increasing spatial resolution improves the agreement between simulations and observations for all variables, with the exception of biomass at the higher resolutions. The GPP agreement improves consistently by 2% with increasing spatial resolution. The canopy height NME improves by 13% from *T42* to *T63*, with a smaller increase of 3% for *T85* compared to *T63*. For the tree cover the improvements going from *T42* to *T63* are appreciable at 14% and 9% with and without the LUC, respectively; however more modest improvements of 6% and 1% (with and without LUC) are seen going from *T63* to *T85*. Biomass is more realistic going from *T42* to *T63* with improvements of 9% (with LUC) and 6% (without LUC) but agreement does not improve notably going from *T63* to *T85*."

This suggests the problem is actually climate biases, and that EMAC should be improved to get a better representation of biomass. The authors may be putting this forward as a pragmatic way of getting the right carbon balance within an atmosphere model which is much harder to fix. If this is the case, then please say so in the text.
Yes, we agree that problem here is in part climate biases, however we were not proposing a pragmatic way of improving the carbon balance in the atmospheric model.

Also, how might the simplistic turn over rates be developed?

Good point. Currently standing wood does not turnover in the LPJ-GUESS . And whilst fine root and leaves turnover rates could be altered, this will not significantly effect standing biomass (which is dominated by wood). We therefore remove the reference to turnover rates.

Reviewers reference:

Avitabile, V., Herold, M. and Heuvelink, G. B. M.: An integrated pan-tropical biomass map using multiple reference datasets, Glob. Chang. Biol. [online] Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13139, 2016.

Bond, W. J., Woodward, F. I. and Midgley, G. F.: The global distribution of ecosystems in a world without fire, New Phytol., 165(2), 525–537, 2005.

Burton, C., Betts, R., Cardoso, M., Feldpausch, R. T., Harper, A., Jones, C. D., Kelley, D. I., Robertson, E. and Wiltshire, A.: Representation of fire, land-use change and vegetation dynamics in the Joint UK Land Environment Simulator vn4.9 (JULES), , doi:10.5194/gmd-12-179-2019, 2019.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model, J. Hydrol., 286(1), 249–270, 2004.

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., Archibald, S., Mouillot, F., Arnold, S. R., Artaxo, P., Bachelet, D., Ciais, P., Forrest, M., Friedlingstein, P., Hickler, T., Kaplan, J. O., Kloster, S., Knorr, W., Laslop, G., Li, F., Melton, J. R., Meyn, A., Sitch, S., Spessa, A., van der Werf, G. R., Voulgarakis, A. and Yue, C.: The status and challenge of global fire modelling, Biogeosciences, 13(11), 3359–3375, 2016.

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 Biogeochem. Cycles, 10(4), 693–709, 1996.

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., Willis, K. O. and Others: A comprehensive benchmarking system for evaluating global vegetation models, Biogeosciences, 10(5), 3313–3340, 2013.

Kloster, S. and Lasslop, G.: Historical and future fire occurrence (1850 to 2100) simulated in CMIP5 Earth System Models, Glob. Planet. Change, 150, 58–69, 2017.

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., Hill, R., Graham, T., Ridley, J., Blaker, A., Calvert, D., Copsey, D., Ellis, R., Hewitt, H., Hyder, P., Ineson, S., Mulcahy, J., Siahaan, A. and Walton, J.: The Low-Resolution Version of HadGEM3 GC3.1: Development and Evaluation for Global Climate, J. Adv. Model. Earth Syst., 10(11), 2865–2888, 2018.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R. and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83(1), 214–231, 2002.

Oak Ridge National Laboratory: Olson's Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: An Updated Database Using the GLC2000 Land Cover Product, [online] Available from: https://cdiac.ess-dive.lbl.gov/epubs/ndp/ndp017/ndp017b.html (Accessed 26 April 2019), n.d.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P. and and Kassem, K. R.: 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, 933–938, 2001.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Li, F., Mangeon, S., Yue, C., Arora, V. K. and Others: The Fire Modeling Intercomparison Project (FireMIP), phase 1: Experimental and analytical protocols, Geoscientific Model Development, 20, 1175–1197, 2017.

Including vegetation dynamics in an atmospheric chemistry-enabled GCM: Linking LPJ-GUESS (v4.0) with EMAC modelling system (v2.53)

Matthew Forrest[1,2], Holger Tost[3], Jos Lelieveld[2,4], and Thomas Hickler[1,5]

[1]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany
[2]Atmospheric Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany
[3]Institute for Atmospheric Physics, Johannes Gutenberg University Mainz, Mainz, Germany
[4]Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus
[5]Institute for Physical Geography, Goethe University, Frankfurt am Main, Germany

Correspondence to: Matthew Forrest (matthew.forrest@senckenberg.de)

Abstract.

Central to the development of Earth System Models (ESMs) has been the coupling of previously separate model types, such as ocean, atmospheric and vegetation models, to ~~provide~~ address interactive feedbacks between the system components. A modelling framework which combines a detailed representation of these components, including vegetation and other land surface processes, enables the study of land-atmosphere feedbacks under global change. ~~This includes the methane cycle and lifetime and the atmospheric chemistry of reduced carbon; fire effects and feedbacks; future nitrogen deposition rates and fertilisation scenarios; ozone damage to plants; and the contribution of biogenic volatile organic compounds to aerosol load and, via cloud condensation nuclei activation, to cloud formation (e.g., precipitation cycles).~~

Here we present the initial steps of coupling LPJ-GUESS, a dynamic global vegetation model, to the atmospheric chemistry enabled atmosphere-ocean general circulation model EMAC. The LPJ-GUESS framework includes a comparatively detailed individual based model of vegetation dynamics ~~, a~~ including a nitrogen cycle. Although not enabled here, the model framework also includes crop and managed-land scheme, a ~~nitrogen cycle~~ representation of arctic methane and permafrost, and a choice of fire models; and hence represents many important terrestrial biosphere processes and provides a wide range of prognostic trace gas emissions from vegetation, soil and fire. ~~When development is complete, these trace gas emissions will form key inputs to the state-of-art atmospheric chemistry representations in EMAC allowing for bi-directional chemical interactions of the surface with the atmosphere. Then the full model will become a powerful tool for investigating land-atmosphere interactions. Initial results show that the~~

We evaluated a one-way, on-line ~~coupling~~ coupled model configuration (with climate variable being passed from EMAC to LPJ-GUESS ~~gives a reasonable~~ but no return information flow) by conducting simulations at three spatial resolution (T42, T63 and T85). These were compared to an expert derived map of potential natural vegetation and four global gridded data products: tree cover, biomass, canopy height and gross primary productivity (GPP). We also applied a post-hoc land use correction to account for human land use. The simulations give a good description of the global potential natural vegetation distribution although there are some regional discrepancies. In particular, at the lower spatial resolutions, a combination of low cold and

low-radiation biases in the growing season of the EMAC climate at high-latitudes causes an underestimation of vegetation extent.

Quantification of the agreement with the gridded datasets using the normalised mean error (NME) averaged over all datasets shows that increasing spatial resolution from T42 to T63 improved the agreement by 10%, and ~~reproduces the broad patterns of biomass, tree cover and canopy height when compared to remote sensing datasets~~ going from T63 to T85 improved agreement by a further 4%. The highest resolution simulation gave an average NME score of 0.67, just 4% worse agreement than an offline LPJ-GUESS simulation using observed climate data (NME = 0.64). However, it should be noted that the offline LPJ-GUESS simulation used a higher spatial resolution which makes the evaluation more rigorous, and that excluding GPP from the datasets (which was anomalously better in the EMAC simulations) gave 10% worse agreement for the EMAC simulation than the offline simulation. Gross primary productivity was best simulated by the coupled simulations, and canopy height the worst. Based on this first evaluation, we conclude that the coupled model provides a suitable means to simulate dynamic vegetation processes into EMAC.

1 Introduction

Simulation models are at the forefront of ~~earth~~ Earth systems research. Historically, such models were initially developed to simulate one component of the ~~earth~~ Earth system in isolation, such as ocean and atmospheric General Circulation Models (GCMs) or Dynamic Global Vegetation Models (DGVMs), with prescribed boundary conditions at the interfaces to other ~~earth~~ Earth system components. However, the interactions between ~~earth~~ Earth system components are dynamic, and representations of feedbacks are necessary to assess the functioning and response of the ~~earth~~ Earth system as a whole. To this end, simulations models have increasingly been coupled to each other to provide dynamic multidirectional fluxes between models, as opposed to prescribing simple non-interacting boundary conditions. This approach has yielded Atmosphere-Ocean General Circulation Models (AOGCMs) which are utilised to understand the dynamics of the physical components of the climate system (Flato et al., 2013).

Interactive carbon cycles and dynamically changing vegetation have been recognised as important processes in the ~~earth~~ Earth system (Cox et al., 2000; Ciais et al., 2013). Consequently, more recent developments have seen AOGCMs extended to include biogeochemical cycles, most often the carbon cycle, to form a new category of model, Earth System Models (ESMs). These state-of-the-art models are the most comprehensive tools for modelling past and future climate change in which biogeochemical feedbacks play an important role, and for studying biosphere-atmosphere feedbacks explicitly (Flato et al., 2013). However, whilst all ESMs by definition have a carbon cycle, not all have truly dynamic vegetation or a nitrogen cycle, fewer still have prognostic representations of fire or of the phosphorous cycle. These processes change vegetation cover and structure in response to changing climate and fire activity, and thus models which do not include them miss key biosphere responses and corresponding feedbacks to the climate system (Wramneby et al., 2010).

To take the first steps towards constructing an ESM with dynamic vegetation, anthropogenic influences and fire, we have combined an atmospheric chemistry-enabled AOGCM, EMAC (Jöckel et al., 2010; Pozzer et al., 2011; Jöckel et al., 2016), with a DGVM, LPJ-GUESS (Smith et al., 2001; Sitch et al., 2003; Smith et al., 2014), in a single modelling framework. LPJ-GUESS ~~is a state-of-the-art DGVM which~~ has been widely applied and extensively evaluated, and has, in different model versions, been extended to include many terrestrial processes and used in over 200 ISI-listed publications[1]. At the time of writing LPJ-GUESS is being actively developed and there are on-going efforts to consolidate many previously independent innovations into the main model release. This combination of active development, broad range of included processes and the flexible modelling framework design of the LPJ-GUESS source code makes it a good choice to provide the land surface component of an ESM.

Furthermore, LPJ-GUESS has already been used both a global ESM, EC-Earth (Weiss et al., 2014; Alessandri et al., 2017), and a regional ESM, RCA-GUESS (Wramneby et al., 2010; Smith et al., 2011; Zhang et al., 2014). In both modelling systems, ~~climate variables and daily soil moisture~~ sub-daily state variables from the atmospheric component and its land surface scheme are aggregated over one simulation day and provided to LPJ-GUESS (Weiss et al., 2014; Smith et al., 2011)~~. In the EC-Earth framework, LPJ-GUESS provides only time-varying leaf area index (LAI) to the atmospheric component which initially only affected physiological resistance (Weiss et al., 2014). This link was recently extended to include effects on albedo, surface roughness length, soil water exploitable by roots and snow shading by vegetation (Alessandri et al., 2017). The land surface scheme in RCA-GUESS splits the gridcell surface into two tiles, one of forest and one herbaceous vegetation, and LPJ-GUESS is used to dynamically adjust the LAI within those tiles and relative fractional coverage of needle-leaved and broad-leaved trees in the forest tile. These LAI and fractional cover values affect albedo, surface roughness and heat fluxes in the land surface scheme (Smith et al., 2011).~~

~~In the work reported here we have adopted a broadly similar approach with regards to forcing LPJ-GUESS with daily aggregated climate fields from the atmospheric model, although daily soil moisture values are calculated by LPJ-GUESS and not the land surface. In the model version described here~~, ~~LPJ-GUESS return LAI, vegetation cover fractions, canopy heights and net primary production to EMAC, which allows dynamic calculation of transpiration (by using the vegetation cover provided by LPJ-GUESS as opposed to prescribed vegetation cover) and of albedo and surface roughness (using newly added parameterisations). However, this information is not (thus far) used by the land surface scheme. In other words, although there is two-way information flow and calculation of land surface properties, the model is only effectively coupled in one direction. Enabling the effect of LPJ-GUESS's dynamic vegetation on the atmosphere (via the land surface scheme) is still under development and will be reported in a future publication (for preliminary results see Tost et al., 2018). The integration of LPJ-GUESS into EMAC is independent of the development of the EC-Earth~~ and ~~RCA-GUESS, but we believe that there are possible synergies in terms of future model development. Furthermore, we consider this parallel development to be complementary in terms of scientific applications, in particular the representation of atmospheric chemistry processes in EMAC allows study of land-atmosphere interactions mediated by trace gas exchanges~~we follow a similar approach here.

[1]see *http://iis4.nateko.lu.se/lpj-guess/LPJ-GUESS_bibliography.pdf* for an up-to-date list of publications featuring LPJ-GUESS

EMAC (ECHAM/MESSy Atmospheric Chemistry)[2] originally combined the ECHAM atmospheric GCM (Roeckner et al., 2006) with the Modular Earth Submodel System (MESSy) (Jöckel et al., 2005) framework and philosophy. The model has since been extended to include ~~state-of-the-art~~ expanded representations of atmospheric chemistry, a coupled ocean model with dynamic sea ice (Pozzer et al., 2011), ocean biogeochemistry (Kern, 2013), an alternative base model for the atmospheric circulation (Baumgaertner et al., 2016), regional downscaling via a two-way coupling (Kerkweg et al., 2018) with the COSMO weather forecast model (Baldauf et al., 2011) and a multitude of processes such as representations for aerosols, aerosol-radiation and aerosol-cloud (Tost, 2017) interactions and many more; all of which are integrated via the MESSy infrastructure.

By bringing together these two ~~state-of-the-art~~ modelling systems, our intent is to produce a fully-featured ESM which benefits from the continuous development of both communities. ~~In addition to the broad range of applications possible for any ESM, the particular strength of EMAC with LPJ-GUESS vegetation will be applications studying interactions and feedbacks at the atmosphere-biosphere boundary, for example: the nitrogen cycle , trace gas emissions from fire, the atmospheric dynamics~~ When development is complete, these trace gas emissions will form key inputs to the atmospheric chemistry representations in EMAC allowing for bi-directional chemical interactions of the surface with the atmosphere. Then the full model will become a powerful tool for investigating land-atmosphere interactions including: the methane cycle and lifetime and the atmospheric chemistry of reduced carbon~~including biogenic volatile organic compound emissions from vegetation and methane from fires, ozone dynamics and the resulting damage to vegetation, and the effects of a wide spectrum of terrestrially emitted trace gases on cloud and aerosol formation anddynamics~~; fire effects and feedbacks; future nitrogen deposition rates and fertilisation scenarios; ozone damage to plants; and the contribution of biogenic volatile organic compounds to aerosol load and, via cloud condensation nuclei activation, to cloud formation (e.g., precipitation cycles).

~~In this manuscript we describe and verify the first steps of our model integration work. Firstly we describe the coupling approach and the technical details of the implementation. Secondly we~~ We plan to follow a step-wise model integration roadmap, whereby the coupling between LPJ-GUESS and EMAC is tightened in well-defined, consecutive steps and processes (such as land use) are included or enabled in a consecutive manner. This will allow us to assess the effects of one model on the other, and the effects of the inclusion of new processes, in a step-wise and logical fashion. For our first step, we have chosen to simulate and evaluate the vegetation ~~state produced by LPJ-GUESS as a result of a one-way coupling whereby~~ produced when LPJ-GUESS is forced ~~on-line by daily climatevalues calculated by EMAC . These results are compared to a suit of remotely-sensed terrestrial biosphere datasets and an expert-derived map of global vegetation cover. Results from the full bidirectional coupling will be reported in a future publication~~by EMAC-simulated climate.

When evaluating the vegetation produced in this configuration, there are potentially three sources of error that may contribute to data-model mismatch: poorly constrained parameters values and inadequate representation of the processes in LPJ-GUESS;

[2] www.messy-interface.org

biases in the climate produced by EMAC (which are expected to have some dependency on the spatial resolution, e.g. see Roeckner et al. (2006)) and missing processes in LPJ-GUESS (predominantly land use). The issue of missing land use was considered in the design of the evaluation method. To disentangle the mismatches resulting from LPJ-GUESS from those resulting EMAC, we consider a 'stand-alone' run of LPJ-GUESS in its standard configuration using observed climate data to assess LPJ-GUESS's implicit biases. To investigate resolution-dependent climate biases in EMAC, simulations with three spatial resolutions were performed and their performance relative to observed data were compared.

2 ~~Model description~~Methods

2.1 ~~The EMAC modelling system~~Model description

2.1.1 The EMAC modelling system

The ECHAM/MESSy Atmospheric Chemistry (EMAC) model is a numerical chemistry and climate simulation system that includes sub-models describing tropospheric and middle atmosphere processes and their interaction with oceans, land and human influences (Jöckel et al., 2010, 2016). The historical starting point for the EMAC model was the ECHAM5 atmospheric model (Roeckner et al., 2006), but the original code has now been fully 'modularised' using the second version of the Modular Earth Submodel System (MESSy2) (Jöckel et al., 2010) including a comprehensive, but highly flexible infrastructure to the point that only the dynamical core and the runtime loop remain from the original code. The physical processes and most of the infrastructure have been split into 'modules' in accordance with the MESSy philosophy whereby such modules can be further developed to improve existing process representations, new modules can be added to represent new processes or alternative process representations, for example parameterised atmospheric convection (Tost et al., 2006), and modules can be selected at run time. EMAC has been extensively used for scientific applications of atmospheric chemistry and chemistry climate interactions from the surface to the mesosphere[3].

2.2 ~~The LPJ-GUESS DGVM (v4.0)~~

2.1.1 The LPJ-GUESS DGVM (v4.0)

The following two paragraphs are modified from a standard LPJ-GUESS model description template which is freely available and copyright free[4]. At its core, LPJ-GUESS (Smith et al., 2001; Sitch et al., 2003; Smith et al., 2014) is a ~~state-of-the-art DGVM featuring a comparatively detailed individual based~~ DGVM featuring an individual-based model of vegetation dynamics. These dynamics are simulated as the emergent outcome of growth and competition for light, space and soil resources among woody plant individuals and a herbaceous understorey in each of a number of replicate patches representing 'random samples' of each simulated locality or grid cell. Multiple patches (in this study 50) are simulated to account for the distribution within

[3] see *http://www.messy-interface.org/* for an up-to-date list of publications featuring MESSy
[4] http://web.nateko.lu.se/lpj-guess/resources.html

a landscape representative of the grid cell as a whole of vegetation stands differing in their histories of disturbance and stand development (succession). The simulated plants are classified into one of a number of plant functional types (PFTs) discriminated by growth form, ~~phenology~~phenological response, photosynthetic pathway (C_3 ~~or~~ and C_4 herbaceous plants), bioclimatic limits for establishment and survival and, for woody PFTs, allometry and life history strategy~~(see Smith et al. (2014) for a description of the standard global PFTs)~~. The standard LPJ-GUESS global PFT set containing 11 plant functional types (needle-and broad-leaved, deciduous and evergreen trees (all of which use C_3 photosynthesis), as well as two types of grass (one C_3 and one C_4) as described in Smith et al. (2014)) was used here. The simulations of this study were carried out in 'cohort mode,' in which, for woody PFTs, cohorts of individuals recruited in the same patch in a given year are identical, and are thus assumed to retain the same size and form as they grow.

Primary production and plant growth follow the approach of LPJ-DGVM (Sitch et al., 2003), and ~~from version 3.0 onwards (we are using v4.0 here) LPJ-GUESS~~ includes an additional nitrogen limitation on photosynthesis (Smith et al., 2014). Canopy fluxes of carbon dioxide and water vapour are calculated by a coupled photosynthesis and stomatal conductance scheme based on the approach of BIOME3 (Haxeltine and Prentice, 1996). The net primary production (NPP) accrued by an average individual plant each simulation year is allocated to leaves, fine roots and, for woody PFTs, sapwood, following a set of prescribed allometric relationships for each PFT, resulting in biomass, height and diameter growth (Sitch et al., 2003). Population dynamics (recruitment and mortality) are represented as stochastic processes, influenced by current resource status, demography and the life history characteristics of each PFT (Hickler et al., 2004, 2012). Forest stand destroying disturbances (such as wind throw and pest attacks) are simulated as a stochastic process, affecting individual patches with an expectation of 0.01 yr^{-1}. All stochastic processes are implemented 'semi-stochastically' using a random number generator with a starting seed. This means that for a fixed starting seed, model runs with the identical settings produce identical results. Litter arising from phenological turnover, mortality and disturbances enters the soil decomposition cycle. ~~As of v3.0, decomposition~~ Decomposition of litter and soil organic matter (SOM) pools follows the CENTURY scheme as described in Smith et al. (2014). Biogenic volatile organic compounds (BVOCs) are emitted from vegetation depending on vegetation type, leaf temperature, atmospheric ~~CO_2~~ CO_2 concentration and carbon assimilation (Arneth et al., 2007a, b). Soil hydrology follows Gerten et al. (2004).

Photosynthesis, respiration and hydrological processes operate on a daily time step and require daily temperature, precipitation and incident short wave radiation. However, monthly climate data may be provided, in which case the model interpolates daily values from the monthly values. In these circumstances, the number of precipitation days in the monthly periods may also be provided to disaggregate total precipitation into distinct rain events. In the case of unmanaged natural vegetation (as simulated here), vegetation dynamics (such as establishment and mortality), disturbance, turnover of plant tissues and turnover between litter pools, and allocation of carbon and nitrogen to plant organs all occur on an annual basis.

~~The~~ Simulation of wildfire is included via the GlobFIRM fire model (Thonicke et al., 2001)~~is included in LPJ-GUESS. It is a simple model~~, which simulates wildfires based on soil moisture (as a proxy for fuel moisture) and a minimum fuel (litter) threshold for burning. ~~Other fire models of greater complexity have been used with LPJ-GUESS: SPITFIRE~~

~~(Lehsten et al., 2009; Thonicke et al., 2010; Rabin et al., 2017), SIMFIRE (Knorr et al., 2016) and SIMFIRE-BLAZE (Rabin et al., 2017). Whilst these models are not currently in the main LPJ-GUESS code version used here, integration efforts are underway and it is is anticipated that they will be available soon in the main LPJ-GUESS version and subsequently in EMAC. Similarly, a representation of tundra, arctic wetlands and permafrost has been developed within a separate branch of LPJ-GUESS (Miller and Smith, 2012) which is now being re-integrated into the main model version. A human land use and agricultural framework (Lindeskog et al., 2013) is included in LPJ-GUESS v4.0 although it is not enabled in this study.~~

2.2 ~~Overview of coupling implementation~~

2.1.1 Overview of coupling implementation

The model coupling strategy employed here was to modify LPJ-GUESS such that it provides its functionality via a new submodel in the MESSy framework. An important design priority was to maintain the integrity of the LPJ-GUESS source code by performing only minimal modifications and additions, in order to facilitate straightforward synchronising with the main LPJ-GUESS trunk version in the future. This approach was successful, with only minor changes made to LPJ-GUESS infrastructure code and no changes to the scientific modules. ~~Three new functions were implemented to exchange information between LPJ-GUESS and MESSy. All these changes were implemented such that LPJ-GUESS can still be compiled outside EMAC and run as a stand-alone model. Within EMAC, the LPJ-GUESS code is compiled into a single library file and the three new functions are called by the new VEG submodel implemented in MESSy. The creation of a single library was necessary (as opposed to direct function calls) because LPJ-GUESS is written in *C++* whereas EMAC is written in *FORTRAN*.~~ For more details see Appendix A.

To provide appropriate climate forcing for LPJ-GUESS, EMAC calculates the daily mean 2 m temperature, daily mean net downwards shortwave radiation and the total daily precipitation at the end of the simulation day and ~~provided~~ provides it to LPJ-GUESS. ~~Atmospheric CO2~~ This is similar to the approach taken by others when coupling LPJ-GUESS to EC-Earth (Weiss et al., 2014; Alessandri et al., 2017) and RCA (Wramneby et al., 2010; Smith et al., 2011; Zhang et al., 2014). However in both these cases, daily soil moisture from the land surface model was also used to drive LPJ-GUESS (which is not done here). Atmospheric CO_2 concentration and nitrogen deposition are also provided on a daily basis from EMAC to LPJ-GUESS. Thus the LPJ-GUESS land-surface state is forced completely by the EMAC atmospheric state and chemical fluxes.

In turn, LPJ-GUESS provides fractional vegetation cover, leaf area index, daily net primary productivity and average height of each PFT to EMAC. ~~Parameterisations for determining albedo and roughness length are implemented in EMAC, however they are not enabled~~ However, these values are not used by EMAC in the simulations presented here. Thus, we are demonstrating only a one-way coupling where the land surface state does not affect the atmospheric state. The boundary conditions for the atmospheric model (in particular the surface energy and water fluxes) come from the pre-existing land surface representation.

~~The coupling strategy is to tighten the coupling between LPJ-GUESS and EMAC in well-defined, consecutive steps and to assess the effects of one model on the other in a step-wise and logical manner. Here we report the effect of EMAC climate on LPJ-GUESS. The next step is to enable the albedo and roughness length schemes and to use the vegetation and forest fractions (which are used in the standard land surface scheme to determine the hydrological fluxes) to form a bidirectional coupling of interactive vegetation and climate. The work is underway (Tost et al., 2018) and will be presented in a future publication.~~

~~Future planned development steps are to enable land use and agriculture in LPJ-GUESS within EMAC, to include a more process-based representation of fire and include the relevant emissions, to fully replace the soil-vegetation part of the hydrological cycle in EMAC with that in LPJ-GUESS and to use LPJ-GUESS to close the land surface energy balance. When completed, these developments will extend the EMAC model into a full Earth system model including atmosphere (ECHAM5) with full chemistry (see Jöckel et al., 2010), vegetation and land surface processes (LPJ-GUESS) and an ocean component (MPIOM) (see Pozzer et al., 2011) with ocean biogeochemistry (see Kern, 2013). The process and exchanges currently included in the modelling framework~~For an overview of the processes and feedbacks enabled in the EMAC configuration used here, as well as ~~planned future additions, are shown in Fig 1.~~ those to be included in future versions, please see Figure 1.

3 ~~Simulation setup~~

2.1 Simulation setup

In the coupled model, the vegetation produced by LPJ-GUESS within EMAC will be directly sensitive to biases in the climate produced by EMAC. It is well-known that these biases are dependent on spatial resolution; see Roeckner et al. (2006) for a study of the biases at different resolutions in ECHAM5 (the GCM upon which EMAC was original based). ~~Whilst it is not within the scope of this study to perform a detailed analysis of the biases in EMAC or their dependence on spatial resolution,~~ Thus the impact of the horizontal spatial resolution of the atmospheric simulation on the vegetation simulation is relevant ~~. To this end we performed two~~ when studying the coupled model setup. To investigate this, we performed three simulations for this evaluation, here denoted *T42*~~and~~ , *T63* and *T85*, which used T42 ~~and~~ (approximately 2.8° × 2.8° grid cell size at the equator), T63 (approximately 1.9° × 1.9° at the equator) and T85 (approximately 1.4° × 1.4° at the equator) spectral resolutions respectively, but were otherwise identical in the considered processes, but use optimal (resolution dependent) "tuning" parameters. As LPJ-GUESS has no inter-gridcell interactions and no processes are gridcell size/spacing dependent, it has no sensitivity to the spatial resolution at which it is run. Thus any improvement in the vegetation produced by the EMAC-coupled simulations at higher resolution can only be due to changes in the EMAC produced climate (i.e. reduced biases). Whilst finer spatial resolution (such as T63 or higher) exhibit lower biases Roeckner et al. (2006) and so are generally preferred wherever possible; coarser resolutions (such as T42) may be used in situations where the large computational cost prohibits finer resolutions, such as long runs transient simulations of paleoclimate, factorial or sensitivity studies, or simulations including detailed atmospheric chemistry calculations.

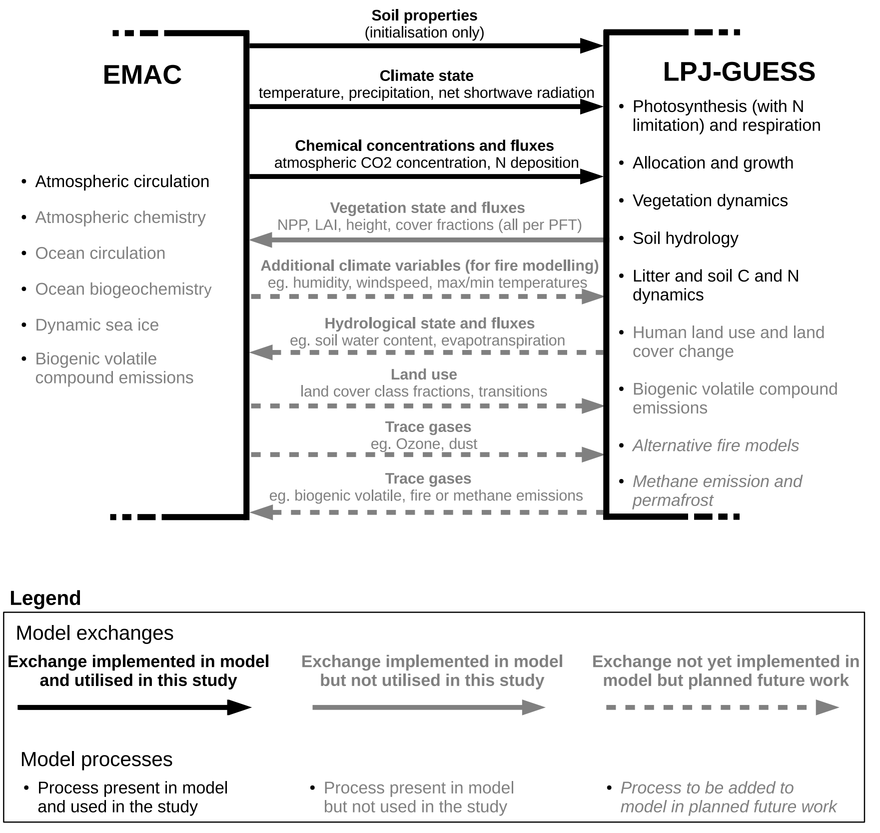


Figure 1. The main processes and exchanges in the coupled model framework. Processes/exchanges with normal black text/black solid arrows are included in the framework and used in the simulations presented here; processes/exchanges with normal grey text/grey solid arrows are included in the framework but not used in the simulations presented here; and processes/exchanges with italic grey text/grey dotted arrows are not included in the framework but planned in future work. All exchanges happen on a daily basis, except for soil properties which happen only during the initialisation phase.

The *T42*, *T63* and *T85* simulations are configured to be equilibrium simulations with boundary conditions corresponding approximately to the late 1990s and early 2000s in order to allow comparisons with the some global datasets from this period (see section 2.1 below). CO_2 was constant and maintained at a level of 367 ppm (corresponding to concentration seen in around the year 1999) throughout the whole simulation. In LPJ-GUESS, nitrogen deposition rates were prescribed using data from Lamarque et al. (2013) for the decade 1990-1990 throughout. The applied EMAC model setup comprised the submodels for radiation (Dietmüller et al., 2016), clouds and convection, surface processes (see Jöckel et al., 2016), and 31 vertical hybrid pressure levels up to 10.0 hPa, representing a typical climate simulation. Note that for reasons of computational burden, the atmospheric chemistry calculations of which EMAC is capable were not activated in these simulations. The model was driven by constant solar (present day) conditions, with prescribed climatological SSTs (sea surface temperatures~~(AMIP II Taylor et al., 2000), including an annual cycle. was constant and maintained at a level of 367 ppm throughout the whole simulation. Throughout both the~~) and SIC (sea ice coverage) from the AMPI2 database (AMIP II Taylor et al., 2000). The climatologies are mean monthly values (so they include the annual cycle) for the years 1995-2000, but do not represent any specific year or include El Nino or La Nina event. Throughout the coupled *T42*, *T63* and *T85* simulations, LPJ-GUESS was driven exclusively by climate variables from EMAC, at no point were external climate datasets used.

As the simulations conducted here utilise only a one-way coupling, EMAC uses its standard land surface scheme which is taken from the ECHAM5 model and is described in detail by Roeckner et al. (2003). Prognostic surface and soil temperatures are calculated with a 5 layer soil model. For the hydrology component, a simple bucket model is assumed, and the water storage capacity is prescribed based on soil type data. A set of land surface data (vegetation ratio, leaf area index, forest ratio, background albedo) has been derived from a global 1 km-resolution dataset for the different horizontal resolutions of the ECHAM5 model (Hagemann, 2002). These data are used to prescribed a climatology of forest fraction (with a constant value) and of vegetation ratio and leaf area index (with a monthly temporal resolution). This prescribed land surface data is used in the model for the calculation of processes such as the interception of precipitation, the snow view in the case of snow-covered surfaces, and for evaporation (bare ground versus vegetated surfaces). Additionally, this data is used in the vertical diffusion scheme and to calculate the grid-mean surface albedo, which depends on a specified background albedo (provided as a constant input data field), a specified snow albedo (function of temperature), the area of the grid cell covered with forest, the snow cover on the ground (function of snow depth and slope of terrain) and the snow cover on the canopy (Roesch et al., 2001).

To aid the interpretation of the EMAC simulations, we also performed an 'offline' LPJ-GUESS simulation using observed climate data from the CRUNCEP bias-corrected, re-analysis dataset (Wei et al., 2014) with a 0.5° spatial resolution. The simulation was performed using exactly the same code and parameter settings as the EMAC *T42*~~and~~ , *T63* and *T85* simulations, but code was compiled as a stand-alone model. The ~~atmopsheric~~ atmospheric CO_2 concentration and nitrogen deposition follow Smith et al. (2014) and the simulation is referred to as the *CRUNCEP* simulation.

In all model simulations a 500 years spin-up phase was used to allow the LPJ-GUESS vegetation to reach approximate equilibrium. The ~~LPJ-GUESS configuration contained 11 plant functional types (needle-and broad-leaved, deciduous and evergreen trees, as well as two types of grass as described in Smith et al. (2014)) and for each gridcell 50 replicate patches were simulated simulated and averaged to account for model stochasticity. Nitrogen deposition rates were prescribed using data from Lamarque et al. (2013) for the decade 1850-1859 throughout the *T42* and *T63* simulations . The degree of nitrogen limitation experienced by the vegetation in the simulations (quantified by the ratio of nitrogen limited to non-nitrogen limited photosythetic rates) is shown in Fig **??**, which shows that pre-industrial nitrogen deposition does not result in additional nitrogen limitation. However, in future work, temporally-appropriate nitrogen deposition data will be used.~~

~~The model simulation duration was dictated by the need to~~ coupled simulations used the online EMAC climate during spin-up~~the LPJ-GUESS vegetation into equilibrium. Here we~~ , and the *CRUNCEP* simulations used the first 30 years (1901-1930) of the CRUNCEP dataset which were detrended and repeated. Simulations followed the standard LPJ-GUESS procedure of starting with ‘bare ground’, ie. no vegetation and no C or N in the soil and litter pools~~, and running for approximately 500 years to allow the vegetation to reach equilibrium~~. Having no plant available N present in the soil at the start of the simulation would inhibit and distort vegetation growth if N limitation was enabled. To overcome this, we ~~follow~~ followed the standard protocol~~and~~ , which is to run LPJ-GUESS for 100 years without N limitation but with normal N deposition to build up the N pools. After 100 years there is sufficient N in the pools, but the vegetation is inconsistent with the desired state as it has been growing without N limitation. Therefore, the vegetation is removed (and the C and N put into the litter pools), and the vegetation is allowed to regrow, this time with N limitation enabled, for a further 400 years. ~~For the *T42* and *T63* simulations, the last 50 years of this 500 year simulation were averaged to produce the plots shown here.~~ At that time, no significant trends in PFT extension and PFT height were obvious, but the vegetation shows interannual variability as expected. ~~In the~~

For the ~~*CRUNCEP*~~*T42*, *T63* and *T85* simulations, an additional 50 years were simulated which were averaged to produce the plots shown here. In the *CRUNCEP* simulation, ~~the first 30 years (1901-1930) were repeated to provide climate data for a 500 years spinup, and then~~ a further 113 years (1901-2013) were simulated using ~~CRUNCEP data~~full CRUNCEP transient time series. The plots presented here show ~~model~~ *CRUNCEP* output averaged over the years 1981-2010.

3 ~~Model evaluation~~

2.1 Model evaluation

~~As~~

Stand-alone LPJ-GUESS has a long history of development and has been evaluated in detail in previous work ~~, for example, net primary production (e.g. Zaehle et al., 2005; Hickler et al., 2006),~~ (some recent examples include modelled potential natural vegetation ~~(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., 2007a),~~ forest stand structure and development ~~(Smith et al., 2001, 2014; Hickler et al., 2004)~~(Smith et al., 2014), global net ecosystem exchange (NEE) variability ~~(Ahlström et al., 2012, 2015) and~~ (Ahlström et al., 2015) and the effect of CO_2 ~~fertilisation experiments (e.g. Hickler et al., 2008; Zaehle e~~ ~~it is beyond the scope of this work to perform a full model evaluation and propose improvements for the dynamic vegetation model. Instead we evaluated~~ fertilisation (e.g. Medlyn et al., 2015)). Here we performed an initial evaluation focused on vegetation state variables relevant to the biophysical coupling between the land and atmosphere in the coupled model setup~~to consider~~ , in order to investigate how LPJ-GUESS responds when EMAC climate is used as the forcing data and to investigate any biases in the vegetation produced.~~For this we used an expert-derived Potential Natural Vegetation (PNV) map and using remotely-sensed data sets of tree cover (Dimiceli et al., 2015), canopy height (Simard et al., 2011) and biomass (Avitabile et al., 2016; Thurner et al., 2014).~~

~~To provide an overall summary metric of data-model agreement across the relevant spatial domain, the Normalised Mean Error (NME) is presented following the prescription and recommendations in Kelley et al. (2013). These summary metrics are not meant to demonstrate a particular level agreement better than some arbitrary threshold. Nor are they mean for strict evaluation or comparison to other models, since here we evaluating only the first milestone of model development and so the model is known to be incomplete (particularly with regard to human land use) and no tuning has been performed. They are included to quantitively evaluate the differences between models runs at difference resolutions and for assessing the effect of human land use via a land cover correction factor (see below).~~

~~It should be noted that the NME is rather different from a coefficient of correlation or a coefficient of determination. It does not attempt to derive a correlation but instead sums the differences between the model and the observation. It can be thought of as quantifying the deviation from the one-to-one line of perfect data-model agreement, rather than the deviation from a line of best fit. This means that is a rather direct and unforgiving metric, since every deviation of the model from the data is penalised (uncertainty is not included) and there is no possibility for the line of best fit to move to compensate for systematic biases. It also means the values are interpreted in the opposite direction to a correlation coefficient; an NME score of zero implies perfect agreement between observation and model, whereas an r^2 of zero would imply no correlation between the two. By the normalisation implicit in the method, using the mean value of the observations in place of the model gives an NME of unity.~~

At this stage of model development we do not seek to precisely simulate the vegetation state of a particular year or exact period. Our atmospheric simulations are not nudged by meteorological data, but rather an unconstrained simulation based on ~~a single year of SSTs~~ climatological SSTs and SIC, so they do not correspond to a particular calendar period. Furthermore we prescribe a fixed atmospheric CO_2 concentration. ~~Instead, our goal with this evaluation is to perform steady state simulations where the climate and forcing are constant and correspond approximately to conditions in the recent past. Thus, after 500 years of simulation, we can compare the equilibrium vegetation to satellite products based on observations in the early 2000s.~~ Whilst we can't expect perfect agreement since (among other reasons) this is not a full transient simulation, the simulations should be sufficient to check if the model coupling is working as intended, and to gain some insight into biases that may be present when LPJ-GUESS is forced using EMAC climate. Instead, our goal with this evaluation is to perform steady state simulations where the climate and CO_2 forcing are constant and correspond approximately to conditions in the recent past. Thus, after 500 years of simulation, we can compare the equilibrium vegetation to satellite products based on observations in the early 2000s. Furthermore, it should be noted that the tree cover and biomass datasets (see section 2.1.1 below) reflect the biosphere as observed in the previous decade or so, and therefore inherently contains the considerable effect of human land use. This results in a conceptual mismatch between the ~~PNV~~ Potential Natural Vegetation (PNV) as simulated by LPJ-GUESS and the observed biosphere state which is relevant when considering these comparisons. To quantify this effect, NME scores including a land use correction (see Appendix D for details) for these datasets are also included in Section 3.2.

~~Whilst it is not within the scope of this work to evaluate the biases of climate state produced by EMAC , knowledge of these~~ Knowledge of EMAC biases is very useful for disentangling the causes of model-data disagreement in the simulated vegetation.

To this end, we include bias plots of seasonal and annual biases in surface temperature, precipitation and net (plant-available) short wave radiation of the EMAC *T42*~~and~~ , *T63* and *T85* climate with respect to the CRUNCEP bias-corrected, re-analysis climate dataset in Appendix B.

2.2 ~~Biomes~~

2.1.1 Evaluation data sets

~~The simulated Leaf Area Index (LAI) was used to classify the vegetation cover into eight "megabiome" types following Forrest et al. (2015). The broad vegetation categories give an overview of the vegetation~~

To provide a visual assessment of the structure and functioning of the vegetation cover at a level of detail relevant for studying interactions between the land surface and the atmosphere~~. These simulations results were compared~~ , we categorised the simulated vegetation into eight "megabiome" types and compared them to an expert-derived ~~potential natural vegetation (PNV) map (Haxeltine and Prentice, 1996) classified into equivalent categories(Smith et al., 2014; Forrest et al., 2015) which was regridded using a largest area fraction algorithm~~ PNV map with equivalent categories. The classification of the simulated vegetation was based on Leaf Area Index (LAI) following Forrest et al. (2015). The expert-derived PNV data was taken from (Haxeltine and Prentice, 1996), regridded to the spatial resolution of the simulations using a largest area fraction algorithm and then aggregated into the eight megabiome classes (Smith et al., 2014; Forrest et al., 2015). It should be borne in mind that there are various sources of uncertainty affecting the classification of biomes in both the data and the model output, such as the somewhat subjective LAI threshold applied to the model data and the inherently subjective nature of expert classification. However, these uncertainties are to some extent minimised by the choice of broad megabiomes (see Forrest et al. (2015) for further discussion) and so, despite this lack of quantitative rigour, such classifications still provide a useful visual method for comparing vegetation cover.

For quantitative evaluation of the simulated vegetation we chose four vegetation state variables which are informative when evaluating ESMs/DGVMs, particularly with regard to the biophysical coupling and carbon flux between the land surface and the atmosphere: fractional coverage of trees; standing biomass; canopy height; and gross primary productivity (GPP). Fractional is coverage of trees is relevant both for the evaluation of stand-alone DGVMs (as it is result of both overall productivity and vegetation dynamics such as tree-grass competition and disturbance regimes) and for land surface schemes (as forested areas have different biophysical properties as non-forested area). To evaluate tree cover, collection 6 of the MOD44B MODIS tree cover (Dimiceli et al., 2015) was averaged between 2000 and 2015 and aggregated using simple averaging to an intermediate resolution of 0.05 degrees. As the MODIS tree cover layer does not include contributions from vegetation under 5m, we added an additional output variable to LPJ-GUESS which sums tree cover from tree individuals taller than 5m only. This variable was used for the comparison to MODIS tree cover.

Standing biomass is a key state variable in ESM and DGVMs as it is connected to productivity, carbon sequestration, evapotranspiration, vegetation cover, canopy height and other critical processes and variables. As such, it is a useful quantity for evaluating DGVM/ESM performance. We produce a near-global map of standing biomass combined two biomass datasets, one tropical (Avitabile et al., 2016) and one northern temperate and boreal (Thurner et al., 2014). These dataset were aggregated to a common spatial resolution to approximately 25 km resolution using simple gridcell averaging and joined the maps (taking the average where they overlapped). Note that no data (non-forested) pixels in the original Thurner et al. (2014) dataset were set to zero to ensure consistency after the averaging procedure with the Avitabile et al. (2016) data. Furthermore, the Avitabile et al. (2016) datasets has no data for a part of the Sahara desert, Arabian peninsula and southern Australia, so no data values are present there.

Canopy height is highly relevant in a land-atmosphere context as it has a direct effect on atmospheric circulation through surface roughness length. To evaluate simulated canopy height, a 1 km tree canopy height map (Simard et al., 2011) was aggregated to an intermediate 10 km resolution by simple averaging (excluding no data values). For comparison, simulated canopy height is calculated from individual tree height (see Appendix C).

GPP is the critical quantity in earth systems modelling, both in terms of the planetary CO_2 budget and in terms of biosphere functioning. We used the global gridded GPP product of Beer et al. (2010) which upscales eddy flux covariance GPP measurements using gridded climate data and a selection of statistical and machine leaning techniques. This dataset provides an average annual value for the period 1995-2005 at 0.5° spatial resolution (which is a suitable 'intermediate-resolution' from which the data can be regridded to the simulation resolutions).

Taken together, these four quantities/data sets capture many of the key features of vegetation structure and functioning which affect biophysical land-atmosphere exchanges. The data sets were re-gridded from their intermediate resolutions to the simulation resolution using second-order conservative remapping (Jones, 1999; CDO (last), 2018). 'Conservative remapping' was initially developed to ensure that fluxes (such as energy and water) are conserved by the remapping processes (Jones, 1999), and is chosen here ensure that global area-weighted sums are conserved. 'Second-order' refers to a variant of the method which produces a smoother interpolation that 'first-order'(Jones, 1999).

2.1.2 Summary metric

To provide an overall summary metric of data-model agreement across the relevant spatial domain, the Normalised Mean Error (NME) is presented following the prescription and recommendations in Kelley et al. (2013). One point where we deviate from their approach is that we use NME for tree cover only, whereas Kelley et al. (2013) use Manhattan Metric (MM) and Square Chord Difference (SCD) to consider the proportions of tree, low-vegetation and bare cover simultaneously. This is because the land use correction represents deforestation by reducing tree cover in managed lands (see Appendix D), but it does not partition the area that was covered previously by trees into bare or low vegetation cover, which would be necessary for the MM or SCD. We therefore prefer not to benchmark bare or low vegetation cover in this study, and simply apply NME to tree cover.

As NME quantifies the absolute error in the model as compared to the data, the relative difference of the values for two models (compared to the same data set) can be considered the relative improvement of one over the other. For example, if one model yields a score of 0.8, and a second yields a score of 0.6, the second can be said to be 25% better than the first, since $0.6 - 0.8/0.8 = -25\%$, ie. a 25% reduction in absolute error.

It should be noted that the NME is rather different from a coefficient of correlation or a coefficient of determination. It does not attempt to derive a correlation but instead sums the differences between the model and the observation. It can be thought of as quantifying the deviation from the one-to-one line of perfect data-model agreement, rather than the deviation from a line of best fit. This means that is a rather direct and unforgiving metric, since every deviation of the model from the data is penalised (uncertainty is not considered) and there is no possibility for the line of best fit to move to compensate for systematic biases. It also means the values are interpreted in the opposite direction to a correlation coefficient; an NME score of zero implies perfect agreement between observation and model, whereas an r^2 of zero would imply no correlation between the two. By the normalisation implicit in the method, using the mean value of the observations in place of the model gives an NME of 1.

3 Results and Discussion

3.1 Spatial Patterns

The simulations ~~reproduce~~ reproduced the global patterns of vegetation ~~cover well (~~type well (Fig 2), although some regional discrepancies are visible. The most obvious mismatch between ~~all~~ the simulations and the reconstructed megabiomes is the underestimation of the abundance of vegetation (in particular tundra) in the high northern latitudes. ~~The tendency is relatively small in the *CRUNCEP* simulation, but larger for the EMAC simulations. The higher resolution~~This is most apparent for the lowest resolution (*T42*) EMAC simulation but improved with increasing spatial resolution, with the *T63* simulation ~~is better than the~~ being better substantially than *T42*~~, as it shows a greater tundra and boreal forest extent. Therefore, this mismatch can be attributed to~~. The EMAC simulation with the highest spatial resolution (*T85*) showed only a small tendency to underestimate high latitude vegetation, to a similar degree as the offline *CRUNCEP* simulation, indicating that this discrepancy is caused by biases in the EMAC climate at low resolution. Examination of GPP in this area (Fig. 3) confirms this by revealing a broad tendency to underestimate GPP above 50°N in the *T42* simulation. This tendency lessens at higher resolution and is not seen in the offline *CRUNCEP* simulation, and can be explained by a high-latitude growing season low temperature and low plant available radiation bias ~~in the EMAC climate~~ at low resolution (Figs B2 and B3)~~at low resolution, which~~. This is somewhat mitigated at higher resolution as would be expected due to a better representation of the synoptic scale systems in T63 ~~(Roeckner et al., 2006).~~ and T85 (Roeckner et al., 2006). This high-latitude underestimation is also visible when comparing

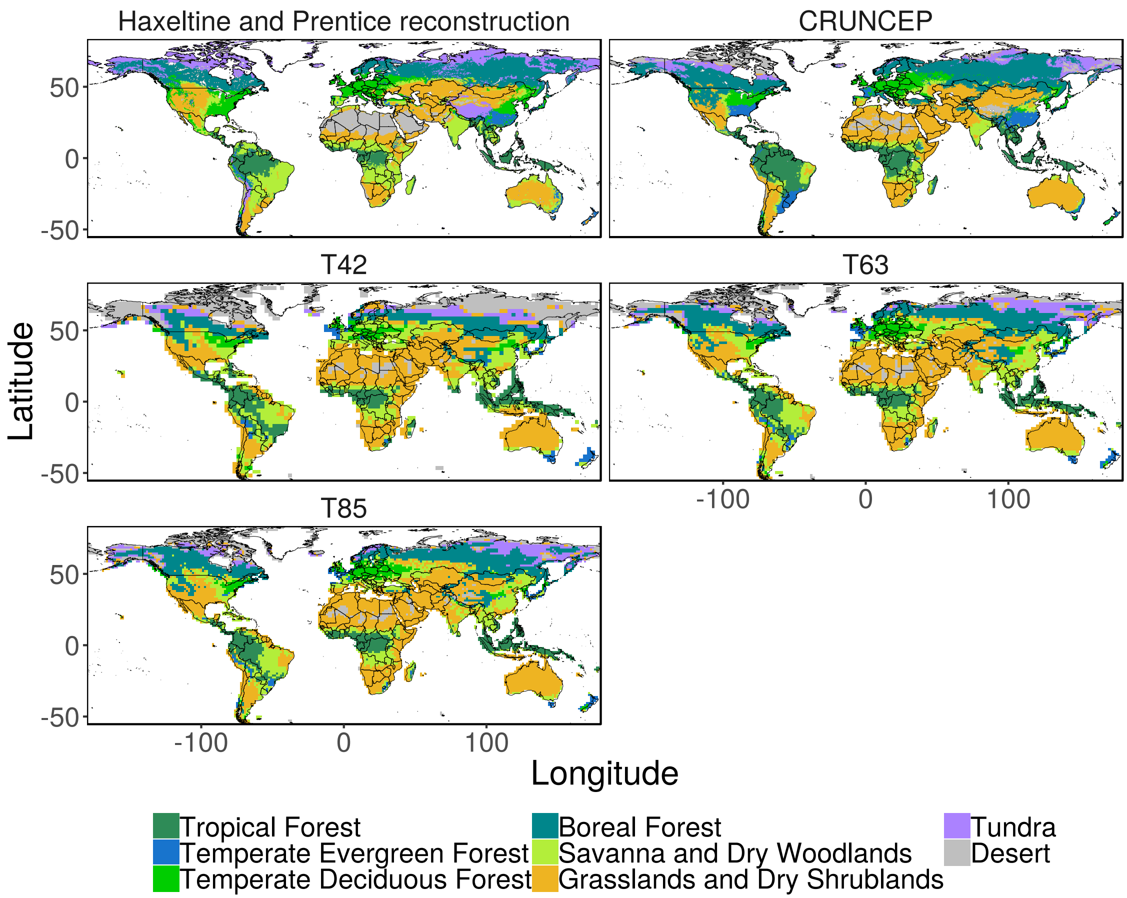


Figure 2. Distribution of PNV megabiomes simulated by LPJ-GUESS within EMAC (*T42*, *T63* and *T85*) and using observed climate data (*CRUNCEP*) compared to an expert-derived PNV map (Haxeltine and Prentice, 1996) following reclassification in (Smith et al., 2014; Forrest et al., 2015).

to observed tree cover (Fig. 4), biomass (Fig. 5) and canopy height (Fig. 6), showing that this issue effects both forested and non-forested vegetation types.

The extent of the temperate forest vegetation zones of the east coasts of the USA and China was underestimated in the ~~simulations using the EMAC climate~~EMAC simulations (Fig 2, 4), which is not seen in the *CRUNCEP* ~~simulations. Inspection of the climate bias plots (Figs B1, B2, B3)shows that this can be~~ simulation. Again, the underlying cause can be identified as insufficient GPP (Fig. 3), this time attributed to a negative bias in precipitation in the southern areas ~~, and~~ (Fig B1), and a negative bias in the plant available radiation in the northern area of coastal China ~~. Large negative precipitation biases reduce the extent of the tropical forest in Indonesia and Papua New Guinea, and in the north-east coast of South America (not seen in~~

~~the *CRUNCEP* simulations). In central Africa and the interior of Brazil, the extent of the tropical forests is also much reduced compared to *CRUNCEP*. However, in this case the reasons are not immediately clear from examination of the bias plots, although some seasonal biases in precipitation and plant available radiation are not clearly apparent~~(Fig. B3). The extent of the Savanna and Dry Woodlands vegetation type ~~is~~ was also under-simulated in Australia and eastern Africa (Fig. 2) as a result of a negative precipitation bias, although this extent ~~is~~ was also not well represented in the *CRUNCEP* simulation. The extent of the Sahara Desert ~~is~~ was also underestimated in the EMAC and *CRUNCEP* simulations, as some areas ~~grow~~ grew sufficient grass cover to be classified as short grasslands. As this is present in the *CRUNCEP* simulation, it is not related to a climate biases, but rather the over-simulation of grasses in very arid regions by LPJ-GUESS.

~~Given that the *T42* and~~ The high productivity of the tropical rainforests was strongly underestimated by the EMAC simulations and to a lesser degree by the ~~*T63*~~*CRUNCEP* ~~results are broadly similar, for brevity of the presentation the subsequent *T42* results will be omitted from the subsequent spatial benchmarks, although the *T42* summary metric results will be tabulated and discussed.~~

~~Distribution of PNV megabiomes simulated by LPJ-GUESS within EMAC (*T42* and *T63*) and using observed climate data (*CRUNCEP*) compared to an expert-derived PNV map.~~

simulation (Fig. 3). This manifests as an underestimation of the extent of the tropical forest vegetation type and tree cover, particularly the east of the Amazon rainforest and in Indonesia and Papua New Guinea (Fig 2 and 4), and as a wide-spread underestimation of biomass and canopy height across the region (Fig 5 and 6). The largest reductions in tropical forest extent coincide with large negative precipitation biases (Fig B1). However, the reasons for the more diffuse underestimation of productivity in the region are not immediately clear from examination of the bias plots, although some seasonal biases in precipitation and plant available radiation are apparent (Figs. B1 and B3). There is also a mild high-temperature bias in some areas (Fig. B2) which may depress productivity by the direct effects of inhibiting photosynthesis and raising plant respiration, and through the indirect effect of exacerbating water-stress conditions due to increasing evapotranspiration. As the *CRUNCEP* simulation also underestimated tropical forest GPP and biomass, there must also be issues with simulating tropical forests in LPJ-GUESS regardless of the climate data used.

The coupled model showed a tendency to overestimate GPP and, to some extent, biomass in the arid continental interiors in central Asia and the central North America. This can be linked to an overestimation of both temperature and precipitation (Figs. B1 and B2). There is also an overestimation (small in magnitude but large as a relative fraction) of biomass in Europe and Eastern China, most likely due to human land use.

3.2 ~~Tree cover~~

~~The collection 6 of the MOD44B MODIS tree cover (Dimiceli et al., 2015) was averaged between 2000 and 2015, interpolated to the simulation resolutions using conservative remapping (Jones, 1999; **?**) and then compared to the simulated tree cover . The~~

Considering tree cover specifically, the combined model produced reasonable global tree cover patterns ~~as would be expected by a state-of-the-art DGVM~~ (Fig. 4a.). However, regional ~~differences are clearly visible~~ discrepancies are starkly visible in the difference plots (Fig. 4b.) which can be attributed to three main sources. The most prominent of these are due to the fact that the simulation is of PNV (ie no human land use processes are included in the simulation) where the observations are of the current state of the planet and therefore include the impact human land use. This conceptual mismatch in the comparison can reasonably account for the large over-estimations of tree cover in Europe, China and temperate North America. The second possible reason for discrepancies in modelled tree cover compared to observed tree cover is ~~climate biases in the EMAC-produced climate. For example, this is~~ poorly simulated productivity in the coupled model. This is most apparent in the underestimation of tree cover on the north-east coast of South America ~~, as already indicated in the biome plots (Fig. 2), which is clearly the result of a large negative precipitation bias in the region (see B1)~~as discussed above. A final source of disagreement is due to the inevitable imperfect process representations in LPJ-GUESS. This is exemplified by the ~~over-estimation of tree cover north of the central African tropical forest~~ encroachment of forest cover into the Sahel in both the EMAC and CRUNCEP forced simulations.

~~Comparison of tree cover from the *T63* (EMAC climate) and *CRUNCEP* (observed climate) simulations with observed tree cover from Dimiceli et al. (2015), a) absolute values and b) the difference between simulations minus observations.~~

3.2 ~~Biomass~~

~~Standing biomass is a key state variable in ESM and DGVMs as it is connected to productivity, carbon sequestration, evapotranspiration, vegetation cover, canopy height and other critical processes and variables which are relevant to vegetation functioning and land-atmosphere exchanges. As such, it is a useful quantity for evaluating DGVM/ESM performance. We combined two biomass datasets, one tropical (Avitabile et al., 2016) and one northern temperate and boreal (Thurner et al., 2014), by aggregating them to approximately 25 km resolution, interpolating them to the simulation resolutions using conservative remapping and then finally joining the maps (taking the average where they overlapped) to produce a nearly global map of standing biomass. Note that no data (non-forested) pixels in the original Thurner et al. (2014) dataset were set to zero to ensure consistency after the averaging procedure with the Avitabile et al. (2016) data.~~

~~Comparison of biomass from the *T63* (EMAC climate) and *CRUNCEP* (observed climate) simulations with observed biomass from Avitabile et al. (2016) and Thurner et al. (2014), a) absolute values and b) the difference between simulations minus observation. Note that neither the Avitabile et al. (2016) nor the Thurner et al. (2014) biomass dataset provide biomass for a band across the Sahara, so no data are plotted there.~~

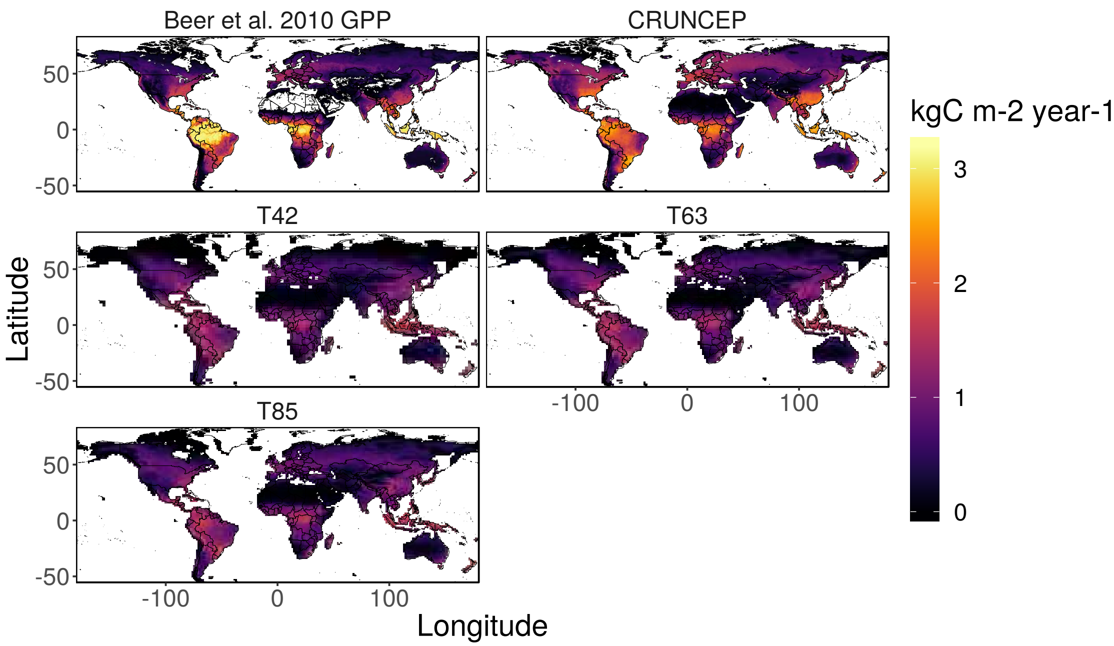
~~The coupled model simulates well the global patterns of biomass (Fig 5), but it does not capture the very high biomass observed in the tropical forests. This underestimation is seen to a lesser extent in the CRUNCEP simulation. The different degrees of underestimation of tropical biomass by the CRUNCEP and EMAC *T63* simulations cannot be explained by the different nitrogen deposition forcing data (**??**), however negative seasonal biases in plant available radiation and precipitation are visible in these regions (Figs B1 and B3). It also underestimates biomass as in the north-east South America and south-east Asia, as would be expected from the biome and tree cover plots (Figs 2 and 4). There is also an over-estimation (small in magnitude but large as a relative fraction) of biomass in Europe and China, most likely~~ This discrepancy cannot be attributed to EMAC climate biases because it also appears in the *CRUNCEP* simulation, neither can it be due to human land use as as the PNV of the area is not forested (Fig.

3.2 ~~Canopy height~~

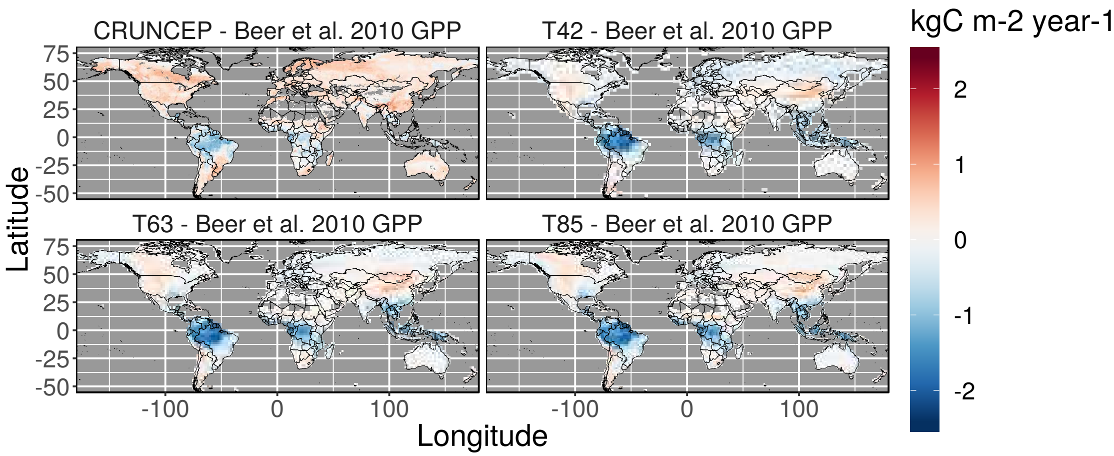
~~In the case of a bidirectional coupling, the simulated canopy height will have a direct effect on atmospheric circulation through roughness length. To evaluate simulated canopy height, a 1 km tree canopy height map (Simard et al., 2011) was first aggregated to 10 km resolution by simple averaging (excluding no data values) and then interpolated to the simulation resolution using conservative remapping. Comparison of these data with simulated canopy height (calculated from individual tree height, see Appendix C) revealed that the coupled model simulates global distributions of canopy height reasonably well (Fig6), but systematically underestimates tree height in highly forested areas. The *CRUNCEP* simulation shows a similar tendency to underestimate canopy height in some regions, indicating that is systemic behaviour in the current LPJ-GUESS parameterisation. In light of the biomass results in Section **??** where biomass is generally underestimated, it may be that the disturbance interval is too frequent, which does not allow sufficient time for biomass and canopy height to build up to realistic levels. Adjusting the disturbance frequency may therefore offer a solution. Another possible cause is the current maximum crown area of trees at 50 m^2 in~~ 2). In such cases the process representations and parameters values in LPJ-GUESS ~~, which is rather low for tropical trees, and may result in an under-weighting of the contribution of mature individuals to canopy height (see Appendix C). In contrast, the simulations tended~~ must be the cause. Another example is the general tendency to overestimate canopy height in arid ~~areas (both *T63*~~ and ~~*CRUNCEP*). This may be attributed~~ semi-arid areas (Fig. 6). Whilst this may be linked to an over-estimation of GPP, it may also be related to the lack of shrub PFTs and/or a low competitiveness of grass PFTs vs tree PFTs ~~, possibly~~ (which in turn may be due to an under-estimation of fire frequency~~. In summary, the pattern of global canopy height is acceptable, but it may be appropriate to adjust the parameterisation in LPJ-GUESS to better reproduce global canopy height in the EMAC framework.~~).

3.2 Summary metrics

The NME scores for ~~both the *T63* and *T42* simulations~~ all simulations and all evaluation variables are presented in Table 1~~(lower is better)~~. In the case of tree cover and biomass, the results are also presented with a land use correction (LUC) factor (see Appendix D). When considering the global NME scores for the coupled model, we see a level of agreement in the ball park of DGVMs used with offline climate data (Kelley et al., 2013), including LPJ-GUESS (Table 1). The fact that the GPP scores for the coupled model are actually better than stand-alone LPJ-GUESS can be dismissed as either an artefact of the coarser spatial resolution of the EMAC simulations or a 'fortuitous cancellation of errors' between the EMAC-produced climate and the LPJ-GUESS parameterisation. It should be noted that in this work we are evaluating only the first milestone of model development, and at this point the model is known to be incomplete (particularly with regard to human land use) and no tuning has been performed to either model component. As such, these summary metrics are not meant to demonstrate a particular level of agreement better than some arbitrary threshold, but rather they are included to quantitatively evaluate the differences between models runs at difference resolutions and to assess the effect of human land use via a land cover correction factor. They

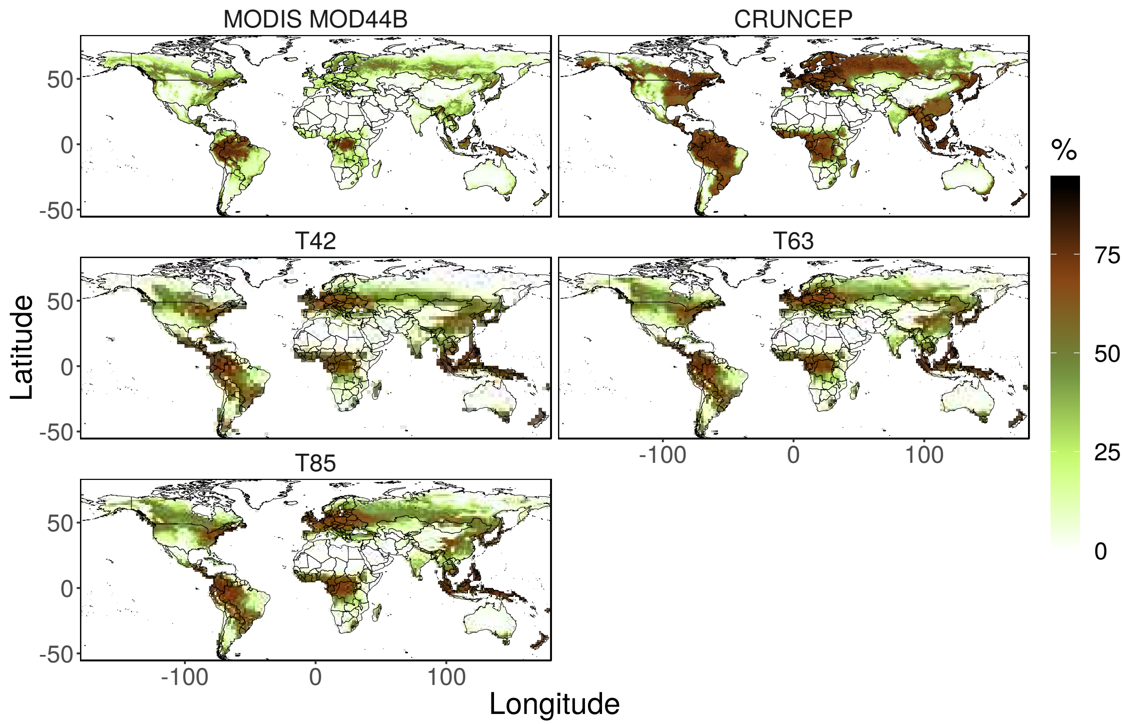


(a)

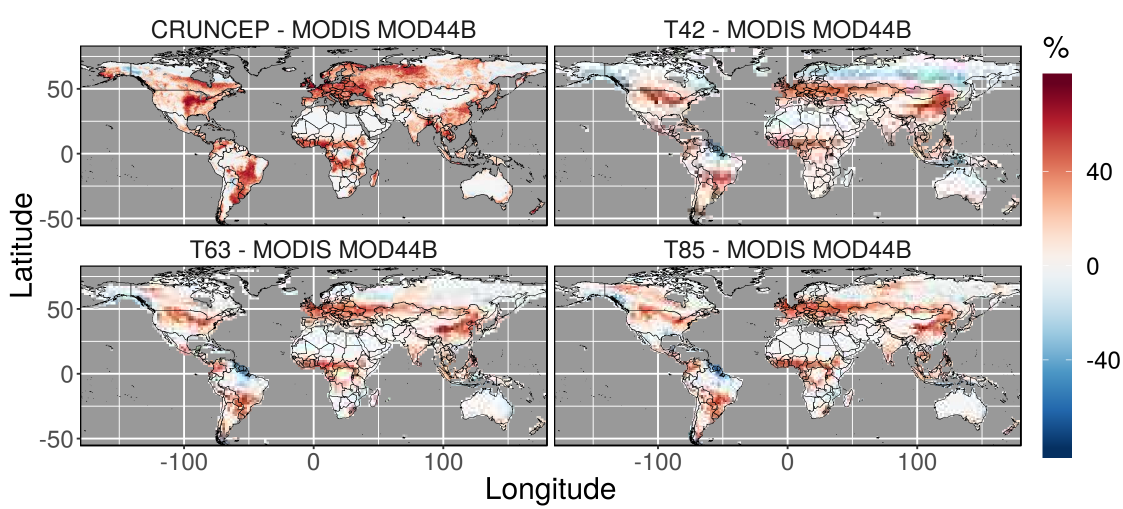


(b)

Figure 3. Comparison of GPP from the *T42*, *T63* and *T85* (EMAC climate) and *CRUNCEP* (observed climate) simulations with the gridded GPP product from Beer et al. (2010), a) absolute values and b) the difference between simulations minus observations.

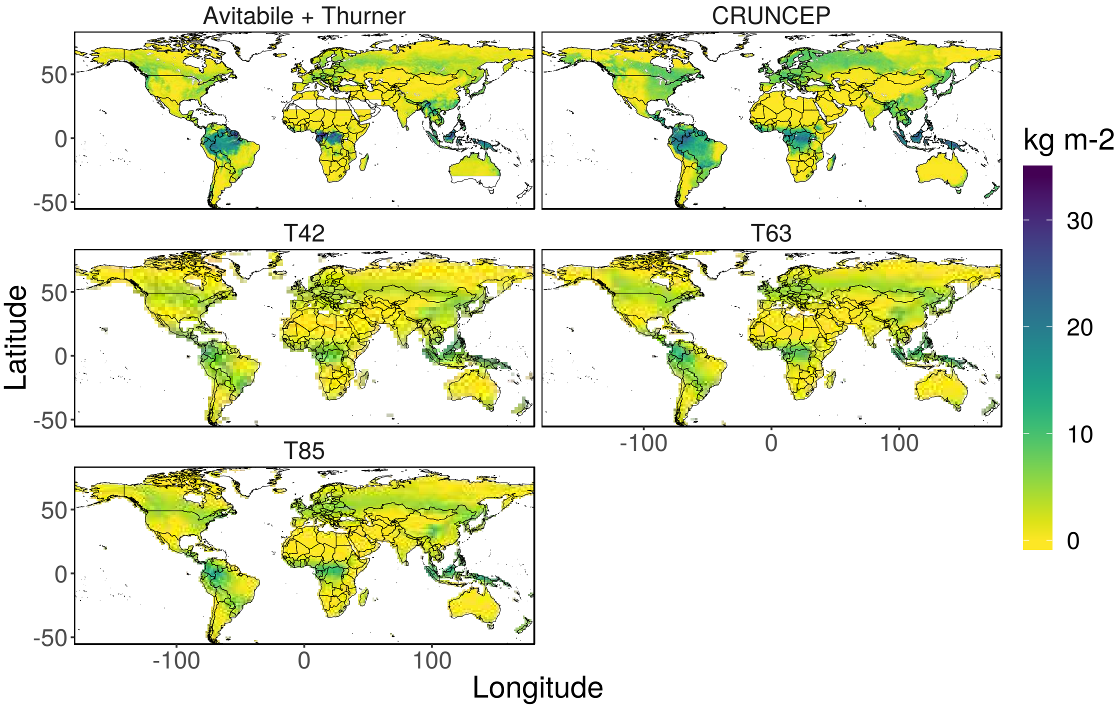


(a)

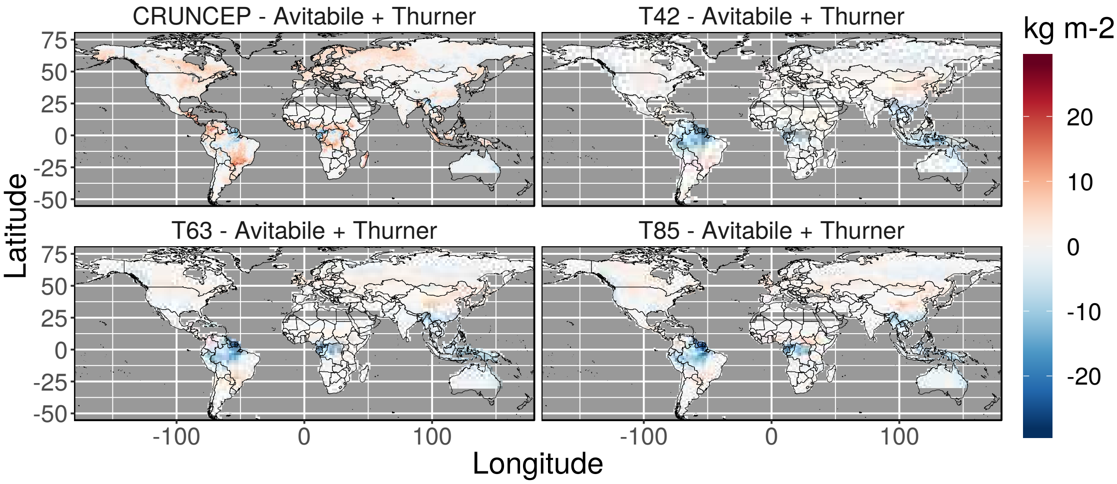


(b)

Figure 4. Comparison of tree cover from the *T42*, *T63* and *T85* (EMAC climate) and *CRUNCEP* (observed climate) simulations with observed tree cover from Dimiceli et al. (2015), a) absolute values and b) the difference between simulations minus observations.

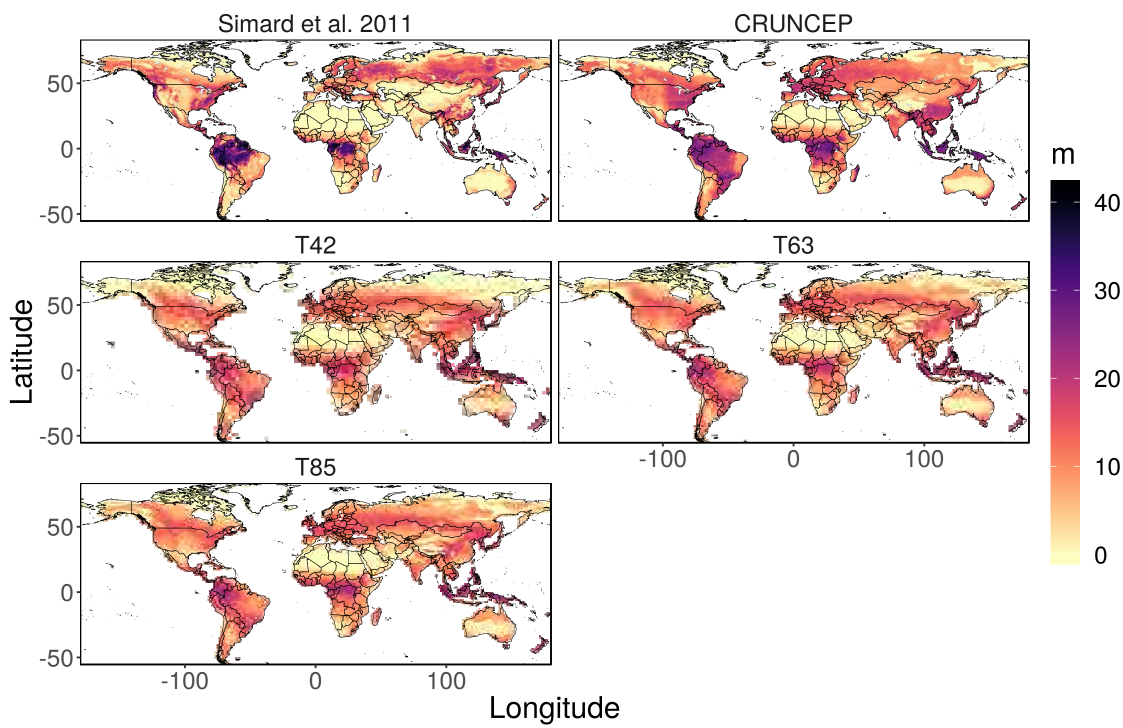


(a)

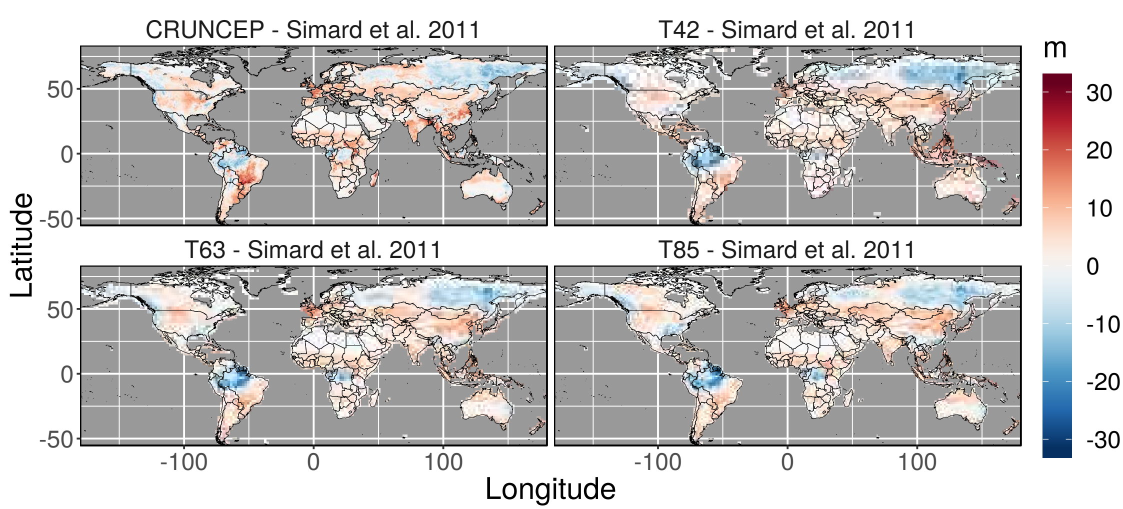


(b)

Figure 5. Comparison of ~~canopy height~~ biomass from the *T42*, *T63* and *T85* (EMAC climate) and ~~CRUNCEP~~ *CRUNCEP* (observed climate) simulations with observed ~~canopy height~~ biomass from ~~Simard et al. (2011)~~Avitabile et al. (2016) and Thurner et al. (2014), a) absolute values and b) the difference between simulations minus observation. Note that neither the Avitabile et al. (2016) nor the Thurner et al. (2014) biomass dataset provide biomass for a band across the Sahara desert and Arabian peninsula and in southern Australia, so no data are plotted there.



(a)



(b)

Figure 6. Comparison of canopy height from the *T42*, *T63* and *T85* (EMAC climate) and *CRUNCEP* (observed climate) simulations with observed canopy height from Simard et al. (2011), a) absolute values and b) the difference between simulations minus observation.

Table 1. NME scores for the vegetation produced by the *T42*, *T63*, *T85* and *CRUNCEP* simulations compared to four gridded global datasets both with and without a LUC (Land Use Correction) where applicable. Note that lower scores imply better agreement between simulation and observation.

	without LUC				with LUC			
Dataset	*T42*	*T63*	*T85*	*CRUNCEP*	*T42*	*T63*	*T85*	*CRUNCEP*
Tree cover (Dimiceli et al., 2015)	0.93	0.85	0.84	1.1	0.78	0.67	0.63	0.62
Biomass (Avitabile et al., 2016; Thurner et al., 2014)	0.72	0.68	0.68	0.76	0.75	0.68	0.66	0.56
Canopy height (Simard et al., 2011)	n/a	n/a	n/a	n/a	1.0	0.87	0.84	0.77
GPP (Beer et al., 2010)	0.55	0.54	0.53	0.60	n/a	n/a	n/a	n/a

The canopy height data was produced in such a way that no land use correction is necessary, and the land use cannot be meaningfully applied to the modelled GPP.

also give a first overview of how the model simulates key features of the vegetated land surface and quantitatively indicate (on a normalised scale) which properties are well simulated and which may require additional tuning in future work. In summary, and especially in light of the caveats above, these good NME scores give confidence that LPJ-GUESS is a suitable choice for coupling to EMAC; that the implementation is working correctly; and that on global scale, LPJ-GUESS is not critically sensitive to biases in the climate produced by EMAC.

Applying the LUC has a marked improvement on the ~~NME scores for tree cover~~tree cover NME scores (in terms of percentage reduction of error: 16% for *T42*, 22% for *T63*, 33% for *T85* and 43% for *CRUNCEP*), implying that ~~some of the discrepancies~~ much of the discrepancy seen between simulation and observation apparent in Fig 4 ~~and 5 are~~ is indeed due to human land use ~~. This also suggests~~ as expected. Whilst this could be expected, it does demonstrate that a large part of the mismatch between the tree cover simulated by the coupled model and the observed tree cover is due to human land use which is not present in the current model. Indeed it highlights the fact that applying a land use scheme or correction will be important when enabling feedbacks from the land surface to the atmosphere in the future~~. For biomass the results are not so clear cut, as including~~ , particularly as tree cover plays a direct role in the determining the biophysical properties of the land surface as seen by the atmospheric model. Whilst applying the land use ~~corrections worsens agreement , particularly for~~ correction does improve biomass agreement in the *CRUNCEP* simulation (by 26%) and the *T85* simulation (by only 3%) , it leaves the agreement in *T63* simulations essentially unchanged 0% change) and actually it worsens agreement in the *T42* simulation by 4%. This can be understood in the context of Fig 5, which shows that the combined model underestimates biomass (particularly at low resolution), and so ~~it can be expected that~~ further reducing the biomass (through the land use correction) will worsen agreement. This indicates the importance of including land use effects in a consistent and realistic way in the coupled model, and that improved simulation of biomass is also critical due its status as a key state variable in the land surface representation. In particular, the average global patch-destroying disturbance rate of 0.01 ~~yr-1~~ yr^{-1} could be re-evaluated and rather simplistic mortality ~~and tissue turnover rates~~ could be further developed in LPJ-GUESS.

For the coupled simulations, increasing spatial resolution ~~also improves~~ improved the agreement between simulations and ~~observation. This indicates that increased spatial resolutionimproves the representation of~~ observations for all variables, with the exception of biomass at the higher resolutions. The GPP agreement improved consistently by 2% with increasing spatial resolution. The canopy height NME improved by 13% from *T42* to *T63*, with a smaller increase of 3% for *T85* compared to *T63*. For the tree cover the improvements going from *T42* to *T63* are appreciable at 14% and 9% with and without the LUC, respectively; however more modest improvements of 6% and 1% (with and without LUC) are seen going from *T63* to *T85*. Biomass is more realistic going from *T42* to *T63*, with improvements of 9% (with LUC) and 6% (without LUC), but agreement did not improve notably going from *T63* to *T85*. Averaging over all the data sets gives an improvement of 10% when going from *T42* to *T63* spatial resolution (with the land use correction applied where appropriate). Going from *T63* to *T85* yields a smaller average improvement of 4%. Whilst this is a decent improvement which may be important in some applications (particularly in the ~~climate in EMAC which in turn tangibly improves the vegetation simulated by~~ case of the fully coupled model runs), the law of diminishing returns clearly applies, so going to higher resolutions than T85 may not be worth the additional computation burden. As LPJ-GUESS ~~. This is particularly noteworthy as~~ processes are completely independent of spatial resolution, these improvements must be coming from the EMAC-simulated climate. It is also worth considering that these gains are particularly noteworthy because conducting benchmarking at higher resolutions is more rigorous and would generally be expected to result in lower benchmark scores. This can be understood mathematically as a consequence of a larger degree of spatial aggregation (of both the evaluation data and the model input data) at coarser resolutions leading to more homogenised values and therefore more agreement. While the result that the resolution of the atmospheric simulation has such a significant effect on the vegetation is not surprising in itself, it does highlight that when considering the bidirectionally coupled model with dynamic (as opposed to prescribed) vegetation, thorough investigation must be made of the effect of the resolution of the atmospheric model on both model components, particularly considering feedbacks between them.

~~NME scores for the vegetation profuced by the *T42*, *T63* and *CRUNCEP* simulations compared to three remotely-sensed global datasets both with and without a LUC (Land Use Correction) where applicable. Note that lower scores imply better agreement between simulation and observation.~~ In light of these results, we would recommend against using T42 resolution given the high latitude growing season temperature and radiation bias and it's effect on the vegetation and, potentially, the resulting feedback to the atmosphere (in the case of the fully coupled model). These effects can be mitigated to a large extent by using T63 resolution without incurring too much additional computation cost. Stepping up to T85 resolution or higher may or not be beneficial depending on the details of the simulation and the study.

3.3 Future work

~~without LUC with LUC~~ The work and simulations presented here are only the first milestone on a planned model integration roadmap. Figure 1 shows the various processes and feedbacks to be enabled as part of this roadmap. The next step is to use albedo and roughness length schemes, and the vegetation and forest fractions (which are used in the standard land surface scheme to determine the hydrological fluxes), to form a bidirectional coupling of interactive vegetation and climate. The work is underway and parameterisations for determining albedo and roughness length and the exchange of the relevant variables are already implemented in the EMAC code base (Tost et al., 2018).

~~Dataset *T42* *T63* *CRUNCEP* *T42* *T63* *CRUNCEP*~~

Following this, the next critical step will be to enable land use and agriculture in LPJ-GUESS within EMAC. This will have be beneficial not just in terms of improving the representation of the land surface as a boundary condition to the atmospheric circulation model, but also because model evaluation and benchmarking will become easier to perform and interpret. However, this step will involve a significant amount of development work to modify the LPJ-GUESS code to receive land cover, state transition and management data from EMAC rather than through the existing channels in LPJ-GUESS. In contrast, the calculation of biogenic volatile organic compounds in LPJ-GUESS will be fairly simple as the only additional variables that are required are daily maximum and minimum temperatures (Arneth et al., 2007a, b).

~~Tree cover(Dimiceli et al., 2015) 0.94 0.85 1.1 0.81 0.69 0.62~~

Another further developmental step is to improve the representation of fire and associated emissions in LPJ-GUESS. The GlobFIRM fire model (Thonicke et al., 2001) is included as a module in LPJ-GUESS and is enabled by default. However, GlobFIRM is very a simple model which simulates wildfires based on soil moisture (as a proxy for fuel moisture) and a minimum fuel (litter) threshold for burning. Other fire models of greater complexity have been used with LPJ-GUESS: SPITFIRE (Lehsten et al., 2009; Thonicke et al., 2010; Rabin et al., 2017), SIMFIRE (Knorr et al., 2016) and SIMFIRE-BLAZE (Rabin et al., 2017). Whilst these models are not currently in the main LPJ-GUESS code version used here, efforts to integrate SIMFIRE-BLAZE are underway, and it is anticipated that it will be available soon in the main LPJ-GUESS version and subsequently in EMAC. Similarly, a representation of tundra, arctic wetlands and permafrost has been developed within a separate branch of LPJ-GUESS (Miller and Smith, 2012) which is now being re-integrated into the main model version.

~~Biomass (Avitabile et al., 2016; Thurner et al., 2014) 0.7 0.8 0.76 0.67 0.7 0.56~~

Initially, LPJ-GUESS was developed as a stand-alone DGVM featuring biogeochemical cycling and vegetation dynamics. It was not designed as a land surface scheme and so some physical properties of the vegetation, such as canopy height, were not high priorities during development. Furthermore, many remotely-sensed data sets, such as the canopy height data used here, were not available during the model's initial development and calibration. It is therefore not surprising that in this study we found that GPP was the best-simulated quantity and canopy height was the least well simulated. Given the direct effect of canopy height on the atmosphere via roughness length, it may be appropriate to adjust the parameterisation in LPJ-GUESS to improve the simulation of canopy height. Candidate parameters include PFT-specific coefficients in the allometric equations (Smith et al., 2001; Sitch et al., 2003), which directly control tree height, and the maximum crown area of trees. Maximum crown area for all trees is currently set to 50 m^2 in LPJ-GUESS, which is rather low for tropical trees (see, for example the maximum reported in Seiler et al. (2014)), and may result in an under-weighting of the contribution of mature individuals to canopy height (see Appendix C). A systematic tuning exercise for LPJ-GUESS may yield appreciable improvements in the representation of canopy height and other important aspects of the global LPJ-GUESS vegetation state.

~~Canopy height (Simard et al., 2011) n/a n/a n~~

Longer term and more ambitious goals on the roadmap are to fully replace the soil-vegetation part of the hydrological cycle in EMAC with that of LPJ-GUESS and to use LPJ-GUESS to close the land surface energy balance. Such developments may benefit from synergies with other on-going coupling work in the LPJ-GUESS community. When completed, these developments will extend the EMAC model into a full Earth system model including atmosphere (ECHAM5) with full chemistry (see Jöckel et al., 2010), vegetation and land surface processes (LPJ-GUESS) and an ocean component (MPIOM) (see Pozzer et al., 2011) with ocean biogeochemistry (see Kern, 2013). Furthermore, coupling efforts to COSMO (Baldauf et al., 2011) via the MESSy framework are undergoing. Additionally, linking LPJ-GUESS to ICON/~~a 0.96 0.81 0.77~~ MESSy in a similar way as for EMAC is straightforward.

4 Conclusions

Here we have reported the first steps towards to producing a new atmospheric-chemistry enabled ESM by combining an atmospheric-chemistry enabled AOGCM with a DGVM. The technical coupling work is now complete and has been achieved in a manner which respects both the integrity and philosophy of the two modelling frameworks, and will therefore allow relatively straightforward updates to both components.

Results from one-way coupled simulations (in which climate information generated by EMAC is used to force LPJ-GUESS but no land-surface information is relayed back to EMAC) showed that the vegetation patterns produced from EMAC climate are reasonable on a global scale. However some regional deviations from the observed vegetation are apparent. Some of these are due to the simple fact that in this configuration LPJ-GUESS produces PNV (potential natural vegetation with no human impacts) while the observed vegetation implicitly includes human impact. This effect was confirmed by performing a correction to account for human land use which improved agreement between simulation and observation. Human land use can be included in future model versions by utilising the recently developed crop and managed land module in LPJ-GUESS (Lindeskog et al., 2013), the use of which should mitigate these issues to a large extent.

A second class of deviations is due to biases in the simulated climate, particularly precipitation biases. This is a more difficult problem to solve; improving climate simulations is the subject of much on-going research. However, it is clear that using higher spatial resolution mitigates climate biases which results in tangible improvements in the simulated vegetation. ~~Furthermore~~Based on the three spatial resolutions, we recommend using T63 resolution as a minimum to due climate biases in the high latitudes in the T42 simulation which resulted in insufficient growth of vegetation. However, using dynamically simulated land surface boundary conditions (in this case from LPJ-GUESS) in a bidirectionally coupled model will alter the atmospheric state and therefore the climate biases. This will be the subject of future studies.

Finally, there are some discrepancies arising as an inevitable consequence of the approximations, missing processes and parameter uncertainties inherent in a process-based model such as LPJ-GUESS. These may be reduced by on-going improvements occurring as LPJ-GUESS is further developed and refined. Given the rather rigorous requirements placed on a biosphere model when bidirectionally coupled to an atmospheric model, it may also be necessary to perform some focused model development work with the goal of improving vegetation functioning and structure so that key biophysical quantities (such as albedo and roughness length) are better simulated. ~~However no simulation model is perfect, and some biases and imperfections are inevitable in any modelcomponent~~Of the variables evaluated here, canopy height was found to be the least well-simulated, suggesting that re-tuning tree height in LPJ-GUESS might be an important step to ensure good performance of the fully coupled model.

Whilst further work remains before the full ESM is completed, we have demonstrated that coupling LPJ-GUESS into the EMAC/MESSy modelling framework has been accomplished, and that LPJ-GUESS provides a suitable basis for an improved and dynamic representation of the land surface in EMAC. A future publication will present a two-way model coupling and investigate the effects of the atmosphere. Once the full coupling has been enabled and calibrated, the resulting model will be ~~a unique~~ powerful tool for investigating atmosphere-biosphere interactions. In addition to the broad range of applications possible for any ESM, the particular strength of EMAC with LPJ-GUESS vegetation will be applications studying interactions and feedbacks at the atmosphere-biosphere boundary, for example: the nitrogen cycle, trace gas emissions from fire, the atmospheric dynamics of reduced carbon including biogenic volatile organic compound emissions from vegetation and methane from fires, ozone dynamics and the resulting damage to vegetation, and the effects of a wide spectrum of terrestrially emitted trace gases on cloud and aerosol formation and dynamics.

Code availability. The Modular Earth Submodel System (MESSy) is continuously further developed and applied by a consortium of institutions. The usage of MESSy and access to the source code is licensed to all affiliates of institutions which are members of the MESSy Consortium. Institutions can become a member of the MESSy Consortium by signing the MESSy Memorandum of Understanding. More information can be found on the MESSy Consortium Website (http://www.messy-interface.org). As the MESSy code is only available under license, no DOI is possible for MESSy code versions. However, the code for coupling to LPJ-GUESS used in this manuscript has already been included in the latest official MESSy version (v2.54).

LPJ-GUESS is used and developed world-wide, but development is managed and the code maintained at Department of Physical Geography and Ecosystem Science, Lund University, Sweden. Model code can be made available to collaborators on entering into a collaboration agreement with the acceptance of certain conditions. The MESSy-coupled version of LPJ-GUESS will be maintained as a derivative of LPJ-GUESS. Because access to LPJ-GUESS is also restricted, no DOI can be assigned to LPJ-GUESS versions. The specific code version used here to enable the MESSy coupling the LPJ-GUESS code in EMAC, code is archived on the LPJ-GUESS subversion server with tag "_publications/MESSY_1.0_20180108" in the catalogue "MESSy". For more details and contact information please see the LPJ-GUESS website (http://web.nateko.lu.se/lpj-guess) or contact the corresponding author.

For review purposes, the code used here is available to the editor and reviewers via a password protected link on condition that the code is for review purposes only, it cannot be used for any other purposes and must be deleted afterwards.

Appendix A: Details of coupling implementation

~~The~~ One of the main priorities during the coupling implementation was to change the LPJ-GUESS source code as little as possible. As such, only the following modifications were made to the LPJ-GUESS code:

- Creation of three new functions to be called externally by the MESSy framework to: initialise an LPJ-GUESS simulation (or restart from a saved state if appropriate); perform one day of LPJ-GUESS simulation given one day of EMAC climate data and return the relevant data; and save the LPJ-GUESS state to disk. These key functions encapsulate the interactions between MESSy and LPJ-GUESS.
- Creation of a new input module (an instantiation of the LPJ-GUESS *C++* class `InputModule`) to handle model initialisation in the MESSy framework, and the inclusion of one extra member function of the `InputModule` class (to read the gridlist file) which was implemented as a dummy function in the other LPJ-GUESS input modules.
- Creation of one additional internal function to calculate the daily values to be handed back to EMAC (such as vegetation cover for a particular PFT).
- Inclusion of an additional output module to save model output useful for benchmarking.
- Minor modifications to the standard output module such that the MPI rank number of each process is added to the file output names allowing the output from each process to be stored in the same directory.
- Minor modifications to the standard LPJ-GUESS restart code to allow the MESSy restart cycle number to be added to the names of the state files to be saved or read by LPJ-GUESS.
- Removal of some of code for the LPJ-GUESS real-time visualisations which is incompatible with the MESSy framework.

No changes to the scientific modules were made, and the directory structure and compilation machinery were untouched. Wherever new code conflicted with the standard offline version, a preprocessor directive was used to ensure that the model still compiled in the standard way outside the MESSy framework. Thus the integrity of LPJ-GUESS was maintained so that updates from the LPJ-GUESS trunk version can be applied relatively easily and the code can still be compiled and run offline.

On the MESSy side, the `Makefile` has been modified to compile the complete LPJ-GUESS code into a single library file using `CMake`, which is LPJ-GUESS's native compilation machinery. This was necessary because LPJ-GUESS is written in *C++* whereas EMAC is written in *FORTRAN*. The LPJ-GUESS library is linked to the rest of the EMAC code with the standard linker of EMAC (also including a link to *C++* standard library). LPJ-GUESS provides functionality to new EMAC submodel (VEG) with its individual submodel interface layer (see Jöckel et al., 2005), which is controlled by a namelist and invokes the above mentioned *C++* functions to communicate with LPJ-GUESS.

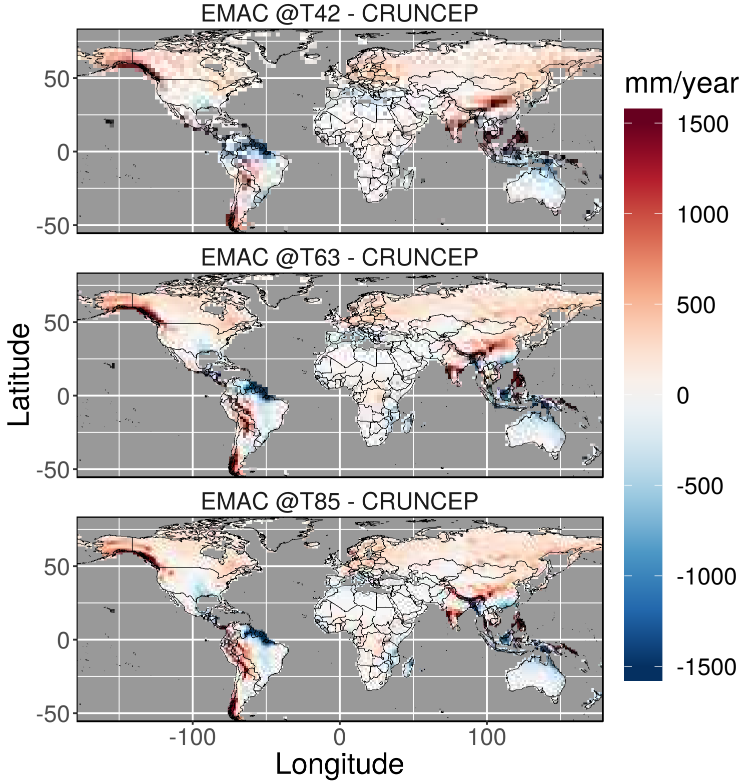
In the initialisation phase, the grid from EMAC is transferred into LPJ-GUESS. Note, that currently there is only a geographic decomposition induced by EMAC, which could lead to some processors not having a single land box and cause idle time for that specific CPU. In future an additional, individual decomposition of the land gridcells to optimise CPU balance is desired, which could make use of the *UniTrans* library developed within the ScalES project[5], which shall also be used for load balanced distribution of chemical gaseous reactions. However, currently the LPJ-GUESS code with its daily timestep consumes very little computing time compared to the climate calculations of EMAC.

In its interface layer, the VEG submodel accumulates the required input fields (daily temperature, precipitation, incoming solar radiation and atmospheric $\mathrm{CO_2}$ concentration) for the vegetation and, depending on the time step length of the LPJ-GUESS code, triggers the call of the LPJ-GUESS routines using the *TIMER*-MESSy interface structure routines (see Jöckel et al., 2005).

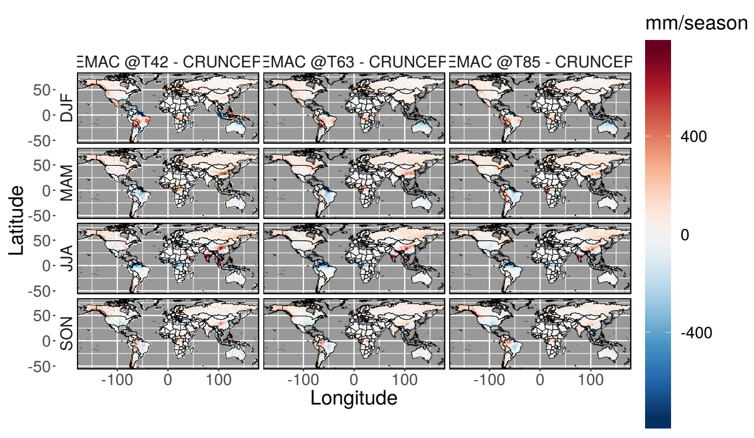
The combined model uses the pre-existing restart facilities of the LPJ-GUESS code, such that when EMAC triggers a restart, a restart is triggered for LPJ-GUESS. When a simulation is continuing from a restart, a flag is sent to the LPJ-GUESS code and the restart files of LPJ-GUESS state are read in allowing a seamless, continuous simulation. This feature may also be used to start a simulation with already well established vegetation from LPJ-GUESS restart (state) files, potentially significant saving significant amounts of CPU time that would otherwise be required to spin up the vegetation (typically the order of 500 simulation years).

[5]https://www.dkrz.de/Klimaforschung/dkrz-und-klimaforschung/infraproj/scales/scales

Appendix B: EMAC climate biases

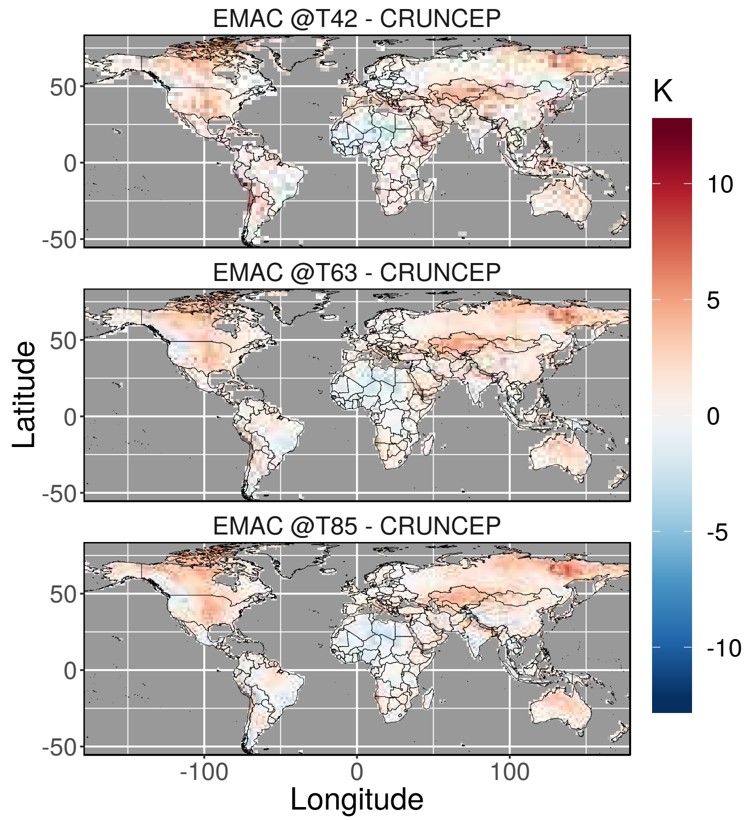


(a)

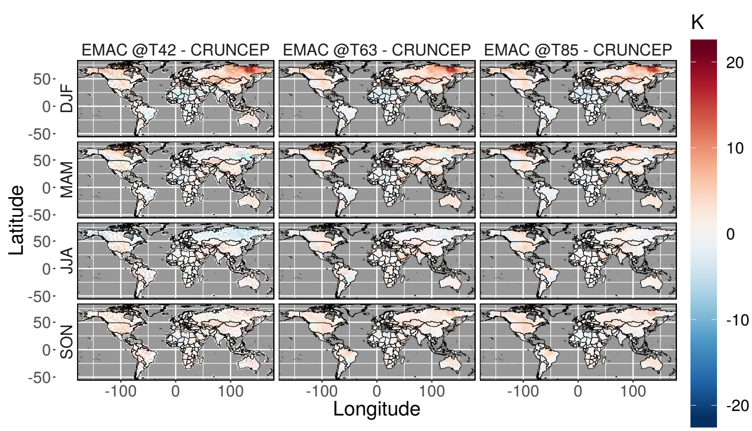


(b)

Figure B1. The a) mean annual precipitation bias and b) mean seasonal precipitation bias between the observed CRUNCEP dataset (1981-2010) and EMAC simulations (last 50 years of simulation). Note that ensure visibility of relatively low precipitations biases, the plotted values are capped at 750 mm/season and 1500 mm/year in the seasonal and annual plots, respectively.

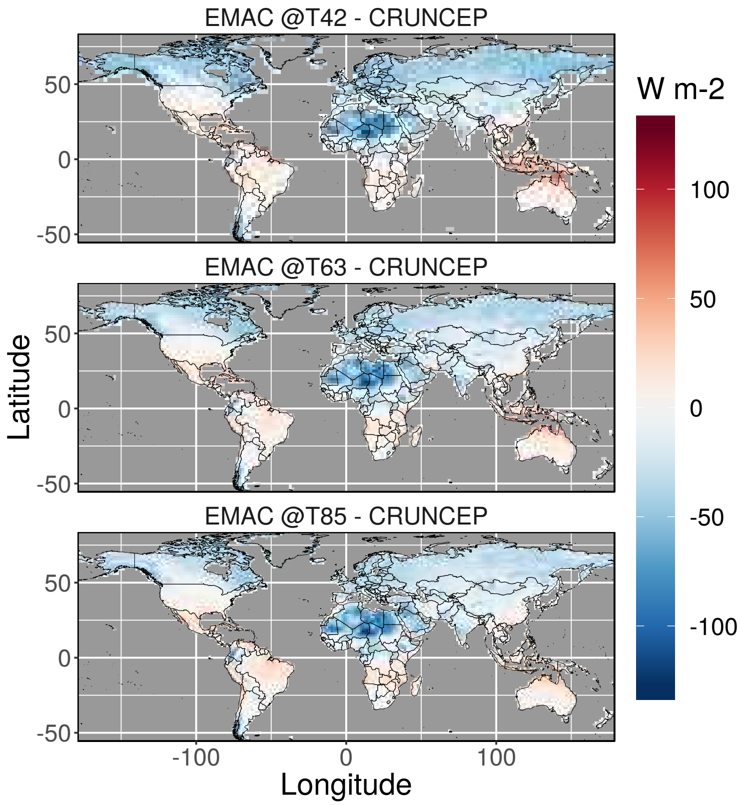


(a)

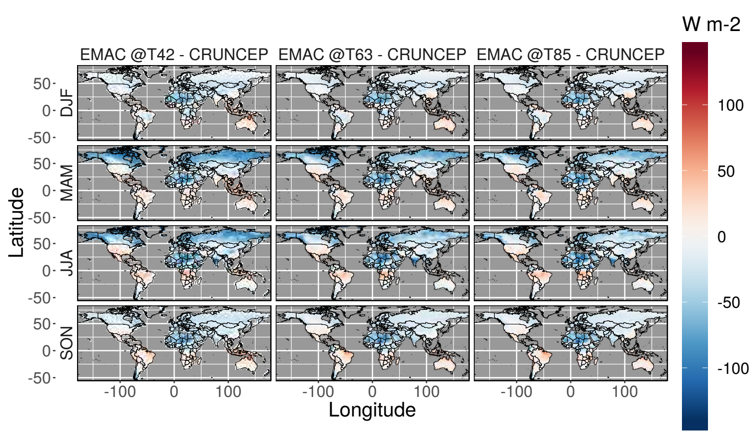


(b)

Figure B2. The a) mean annual temperature bias and b) mean seasonal temperature bias between the observed CRUNCEP dataset (1981-2010) and EMAC simulations (last 50 years of simulation).



(a)



(b)

Figure B3. The a) mean annual net shortwave radiation bias and b) mean seasonal net shortwave radiation bias between the observed CRUNCEP (1981-2010) and EMAC simulations (last 50 years of simulation). Note that these plots compare shows the radiation available for vegetation in the CRUNCEP and EMAC forced LPJ-GUESS simulations, and consequently the gross shortwave radiation has been adjusted by different albedo values. The CRUNCEP radiation has been adjusted using the standard LPJ value of 0.17 applied (temporally and spatially invariant), and the EMAC radiation has been adjusted by the spatially and temporally varying albedo values in the land surface scheme.

Appendix C: ~~Nitrogen limitation~~

~~The ratio of nitrogen limited to nitrogen non-limited photosynthetic rates (weighted by leaf biomass) in all simulation runs.~~

Appendix C: Canopy height calculation

Canopy height of a patch was calculated from individual tree cohort heights by a simple algorithm that attempts to reconstruct top of canopy height as it would be viewed from above, for example by a satellite. It utilises the modelled quantity Foliar Projective Cover (FPC), which is the ground area covered by the crowns of trees of a cohort expressed as a fraction of the patch area. LPJ-GUESS allows limited overlapping of trees and hence the sum of tree cohort FPC can be greater than unity. In this case cohorts are selected in descending order of height until the sum of their FPC reaches 1, i.e. smaller cohorts are assumed to be under the taller cohorts and so do not contribute to top of canopy height. Cohorts smaller than 5 m don't contribute to canopy height as the remotely-sensed dataset does not include canopies lower than 5 m. Having selected the contributing tree cohorts, the canopy height is simply the FPC-weighted sum of the contributing cohort heights.

Appendix D: Land use correction

In order to correct the model output for 'missing' tree cover and biomass due to human land cover modification, a simple correction was derived from the Globcover2009 land cover product (Arino et al., 2012). For each simulated gridcell the fraction of naturally vegetated land pixels from the Globcover2009 product was calculated. This fraction was then used to scale the model outputs of tree cover and biomass to give a simple, first order reduction based on remotely-sensed data. For these purposes, natural vegetated land cover was defined ~~as classeswas defined~~ by the following classes in the Globcover 2009 dataset:

- 40 Closed to open broadleaved evergreen or semi-deciduous forest
- 50 Closed broadleaved deciduous forest
- 60 Open broadleaved deciduous forest/woodland
- 70 Closed needleleaved evergreen forest
- 90 Open needleleaved deciduous or evergreen forest
- 100 Closed to open mixed broadleaved and needleleaved forest
- 110 Mosaic forest or shrubland/ grassland
- 120 Mosaic grassland/forest or shrubland
- 130 Closed to open (broadleaved or needleleaved, evergreen or deciduous) shrubland

– 140 Closed to open herbaceous vegetation (grassland, savannas or lichens/mosses)

Author contributions. HT and MF performed the model coupling. MF performed the simulations and analysis. All authors contributed to the overall model coupling strategy and to the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. Parts of this research were conducted using the supercomputer Mogon and/or advisory services offered by Johannes Gutenberg University Mainz (hpc.uni-mainz.de), which is a member of the AHRP (Alliance for High Performance Computing in Rhineland Palatinate, www.ahrp.info) and the Gauss Alliance e.V. Further development and the main simulations were performed using the LOEWE-CSC supercomputer at the High Performance Computing initiative. The authors gratefully acknowledge the computing time granted on the supercomputer Mogon at Johannes Gutenberg University Mainz (hpc.uni-mainz.de) and the LOEWE-CSC supercomputer at Goethe University Frankfurt (csc.uni-frankfurt.de).
The authors acknowledge the support of the MESSy core development team and are grateful for hints and discussions. Similarly, the authors recognise and appreciate the many improvements to LPJ-GUESS by the LPJ-GUESS development team which made this work possible, and thank the team (particularly Johan Nord) for their support. We also thank and acknowledge Allan Spessa for his contributions at the conception stage of this project.

References

Ahlström, A., Schurgers, G., Arneth, A., and Smith, B.: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections, Environmental Research Letters, 7, 044 008, https://doi.org/10.1088/1748-9326/7/4/044008, http://stacks.iop.org/1748-9326/7/i=4/a=044008, 2012.

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., and Zeng, N.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink, Science, 348, 895–899, https://doi.org/10.1126/science.aaa1668, http://science.sciencemag.org/content/348/6237/895, 2015.

Alessandri, A., Catalano, F., Felice, M. D., Hurk, B. V. D., Reyes, F. D., Boussetta, S., Balsamo, G., and Miller, P. A.: Multi-scale enhancement of climate prediction over land by increasing the model sensitivity to vegetation variability in EC-Earth, Climate Dynamics, 49, 1215–1237, https://doi.org/10.1007/s00382-016-3372-4, https://link.springer.com/article/10.1007/s00382-016-3372-4, 2017.

Arino, O., Perez, J. J. R., Kalogirou, V., Bontemps, S., Defourny, P., and Bogaert, E. V.: Global Land Cover Map for 2009 (GlobCover 2009), PANGAEA, https://doi.org/10.1594/PANGAEA.787668, https://doi.org/10.1594/PANGAEA.787668, 2012.

Arneth, A., Miller, P. A., Scholze, M., Hickler, T., Schurgers, G., Smith, B., and Prentice, I. C.: CO2 inhibition of global terrestrial isoprene emissions: Potential implications for atmospheric chemistry, Geophysical Research Letters, 34, L18 813, https://doi.org/10.1029/2007GL030615, http://onlinelibrary.wiley.com/doi/10.1029/2007GL030615/abstract, 2007a.

Arneth, A., Niinemets, U., Pressley, S., Bäck, J., Hari, P., Karl, T., Noe, S., Prentice, I. C., Serça, D., Hickler, T., Wolf, A., and Smith, B.: Process-based estimates of terrestrial ecosystem isoprene emissions: incorporating the effects of a direct CO2-isoprene interaction, Atmos. Chem. Phys., 7, 31–53, https://doi.org/10.5194/acp-7-31-2007, https://www.atmos-chem-phys.net/7/31/2007/, 2007b.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Ferry, S. J. W., Sunderland, T., Laurin, G. V., Gatti, R. C., Valentini, R., Verbeeck, H., Wijaya, A., and Willcock, S.: An integrated pan-tropical biomass map using multiple reference datasets, Global Change Biology, 22, 1406–1420, https://doi.org/10.1111/gcb.13139, http://onlinelibrary.wiley.com/doi/10.1111/gcb.13139/abstract, 2016.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Monthly Weather Review, 139, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1, https://journals.ametsoc.org/doi/10.1175/MWR-D-10-05013.1, 2011.

Baumgaertner, A. J. G., Jöckel, P., Kerkweg, A., Sander, R., and Tost, H.: Implementation of the Community Earth System Model (CESM) version 1.2.1 as a new base model into version 2.50 of the MESSy framework, Geosci. Model Dev., 9, 125–135, https://doi.org/10.5194/gmd-9-125-2016, https://www.geosci-model-dev.net/9/125/2016/, 2016.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 329, 834–838, https://doi.org/10.1126/science.1184984, https://science.sciencemag.org/content/329/5993/834, 2010.

CDO (last): Climate Data Operators, http://www.mpimet.mpg.de/cdo, 2018.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R. B., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 465–570, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, 408, 184–187, https://doi.org/10.1038/35041539, https://www.nature.com/articles/35041539, 2000.

Dietmüller, S., Jöckel, P., Tost, H., Kunze, M., Gellhorn, C., Brinkop, S., Frömming, C., Ponater, M., Steil, B., Lauer, A., and Hendricks, J.: A new radiation infrastructure for the Modular Earth Submodel System (MESSy, based on version 2.51), Geosci. Model Dev., 9, 2209–2222, https://doi.org/10.5194/gmd-9-2209-2016, https://www.geosci-model-dev.net/9/2209/2016/, 2016.

Dimiceli, C., Carroll, M., Sohlberg, R., Kim, D. H., Kelly, M., and Townshend, J. G. R.: MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org//10.5067/MODIS/MOD44B.006, https://doi.org/10.5067/MODIS/MOD44B.006, 2015.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W. J., Cox, P., Driouech, F., Emori, S., Eyring, V., and Others: Evaluation of Climate Models., in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 741–866, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Forrest, M., Eronen, J. T., Utescher, T., Knorr, G., Stepanek, C., Lohmann, G., and Hickler, T.: Climate-vegetation modelling and fossil plant data suggest low atmospheric CO2 in the late Miocene, Clim. Past, 11, 1701–1732, https://doi.org/10.5194/cp-11-1701-2015, http://www.clim-past.net/11/1701/2015/, 2015.

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, https://doi.org/10.1016/j.jhydrol.2003.09.029, http://www.sciencedirect.com/science/article/pii/S0022169403003901, 2004.

Hagemann, S.: An improved land surface parameter dataset for global and regional climate models., Tech. Rep. 336, Max-Planck-Institut für Meteorologie, Hamburg, https://doi.org/10.17617/2.2344576, 2002.

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, https://doi.org/10.1029/96GB02344, http://onlinelibrary.wiley.com/doi/10.1029/96GB02344/abstract, 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, https://doi.org/10.1890/02-0344, http://onlinelibrary.wiley.com/doi/10.1890/02-0344/abstract, 2004.

Hickler, T., Eklundh, L., Seaquist, J. W., Smith, B., Ardö, J., Olsson, L., Sykes, M. T., and Sjöström, M.: Precipitation controls Sahel greening trend, Geophysical Research Letters, 32, https://doi.org/10.1029/2005GL024370, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL024370, 2005.

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, https://doi.org/10.1111/j.1466-8238.2006.00254.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2006.00254.x/abstract, 2006.

Hickler, T., Smith, B., Prentice, I. C., MjöFors, K., Miller, P., Arneth, A., and Sykes, M. T.: CO_2 fertilization in temperate FACE experiments not representative of boreal and tropical forests, Global Change Biology, 14, 1531–1542, https://doi.org/10.1111/j.1365-2486.2008.01598.x, http://doi.wiley.com/10.1111/j.1365-2486.2008.01598.x, 2008.

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, https://doi.org/10.1111/j.1466-8238.2010.00613.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2010.00613.x/abstract, 2012.

Jones, P. W.: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates, Monthly Weather Review, 127, 2204–2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, https://journals.ametsoc.org/doi/full/10.1175/1520-0493(1999)127%3C2204:FASOCR%3E2.0.CO;2, 1999.

Jöckel, P., Sander, R., Kerkweg, A., Tost, H., and Lelieveld, J.: Technical Note: The Modular Earth Submodel System (MESSy) - a new approach towards Earth System Modeling, Atmos. Chem. Phys., 5, 433–444, https://doi.org/10.5194/acp-5-433-2005, https://www.atmos-chem-phys.net/5/433/2005/, 2005.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geosci. Model Dev., 3, 717–752, https://doi.org/10.5194/gmd-3-717-2010, https://www.geosci-model-dev.net/3/717/2010/, 2010.

Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51, Geosci. Model Dev., 9, 1153–1200, https://doi.org/10.5194/gmd-9-1153-2016, https://www.geosci-model-dev.net/9/1153/2016/, 2016.

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., and Willis, K. O.: A comprehensive benchmarking system for evaluating global vegetation models, Biogeosciences, 10, 3313–3340, https://doi.org/10.5194/bg-10-3313-2013, http://www.biogeosciences.net/10/3313/2013/, 2013.

Kerkweg, A., Hofmann, C., Jöckel, P., Mertens, M., and Pante, G.: The on-line coupled atmospheric chemistry model system MECO(n) – Part 5: Expanding the Multi-Model-Driver (MMD v2.0) for 2-way data exchange including data interpolation via GRID (v1.0), Geosci. Model Dev., 11, 1059–1076, https://doi.org/10.5194/gmd-11-1059-2018, https://www.geosci-model-dev.net/11/1059/2018/, 2018.

Kern, B.: Chemical interaction between ocean and atmosphere, http://ubm.opus.hbz-nrw.de/volltexte/2014/3732, 2013.

Knorr, W., Jiang, L., and Arneth, A.: Climate, CO2 and human population impacts on global wildfire emissions, Biogeosciences, 13, 267–282, https://doi.org/10.5194/bg-13-267-2016, https://www.biogeosciences.net/13/267/2016/, 2016.

Lamarque, J.-F., Dentener, F., McConnell, J., Ro, C.-U., Shaw, M., Vet, R., Bergmann, D., Cameron-Smith, P., Dalsoren, S., Doherty, R., Faluvegi, G., Ghan, S. J., Josse, B., Lee, Y. H., MacKenzie, I. A., Plummer, D., Shindell, D. T., Skeie, R. B., Stevenson, D. S., Strode, S., Zeng, G., Curran, M., Dahl-Jensen, D., Das, S., Fritzsche, D., and Nolan, M.: Multi-model mean nitrogen and sulfur deposition from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): evaluation of historical and projected future changes, Atmos. Chem. Phys., 13, 7997–8018, https://doi.org/10.5194/acp-13-7997-2013, https://www.atmos-chem-phys.net/13/7997/2013/, 2013.

Lehsten, V., Tansey, K., Balzter, H., Thonicke, K., Spessa, A., Weber, U., Smith, B., and Arneth, A.: Estimating carbon emissions from African wildfires, Biogeosciences, 6, 349–360, https://doi.org/10.5194/bg-6-349-2009, https://www.biogeosciences.net/6/349/2009/, 2009.

Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., and Smith, B.: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa, Earth Syst. Dynam., 4, 385–407, https://doi.org/10.5194/esd-4-385-2013, http://www.earth-syst-dynam.net/4/385/2013/, 2013.

Lucht, W., Prentice, I. C., Myneni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., Bousquet, P., Buermann, W., and Smith, B.: Climatic Control of the High-Latitude Vegetation Greening Trend and Pinatubo Effect, Science, 296, 1687–1689, https://doi.org/10.1126/science.1071828, http://science.sciencemag.org/content/296/5573/1687, 2002.

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., and Norby, R. J.: Using ecosystem experiments to improve vegetation models, Nature Climate Change, 5, 528–534, https://doi.org/10.1038/nclimate2621, https://www.nature.com/articles/nclimate2621, 2015.

Miller, P. A. and Smith, B.: Modelling Tundra Vegetation Response to Recent Arctic Warming, AMBIO, 41, 281–291, https://doi.org/10.1007/s13280-012-0306-1, https://link.springer.com/article/10.1007/s13280-012-0306-1, 2012.

Pozzer, A., Jöckel, P., Kern, B., and Haak, H.: The Atmosphere-Ocean General Circulation Model EMAC-MPIOM, Geosci. Model Dev., 4, 771–784, https://doi.org/10.5194/gmd-4-771-2011, https://www.geosci-model-dev.net/4/771/2011/, 2011.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geosci. Model Dev., 10, 1175–1197, https://doi.org/10.5194/gmd-10-1175-2017, http://www.geosci-model-dev.net/10/1175/2017/, 2017.

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 ECHAM 5. Part I: Model description, Tech. Rep. 349, Max-Planck-Institut für Meteorologie, Hamburg, https://doi.org/10.17617/2.995269, 2003.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of Simulated Climate to Horizontal and Vertical Resolution in the ECHAM5 Atmosphere Model, Journal of Climate, 19, 3771–3791, https://doi.org/10.1175/JCLI3824.1, http://journals.ametsoc.org/doi/full/10.1175/JCLI3824.1, 2006.

Roesch, A., Wild, M., Gilgen, H., and Ohmura, A.: A new snow cover fraction parametrization for the ECHAM4 GCM, Climate Dynamics, 17, 933–946, https://doi.org/10.1007/s003820100153, 2001.

Seiler, C., Hutjes, R. W. A., Kruijt, B., Quispe, J., Añez, S., Arora, V. K., Melton, J. R., Hickler, T., and Kabat, P.: Modeling forest dynamics along climate gradients in Bolivia, Journal of Geophysical Research: Biogeosciences, 119, 758–775, https://doi.org/10.1002/2013JG002509, http://onlinelibrary.wiley.com/doi/10.1002/2013JG002509/abstract, 2014.

Simard, M., Pinto, N., Fisher, J. B., and Baccini, A.: Mapping forest canopy height globally with spaceborne lidar, Journal of Geophysical Research: Biogeosciences, 116, G04 021, https://doi.org/10.1029/2011JG001708, http://onlinelibrary.wiley.com/doi/10.1029/2011JG001708/abstract, 2011.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., 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, https://doi.org/10.1046/j.1365-2486.2003.00569.x, http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2486.2003.00569.x/abstract, 2003.

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, https://doi.org/10.1046/j.1466-822X.2001.t01-1-00256.x, http://onlinelibrary.wiley.com/doi/10.1046/j.1466-822X.2001.t01-1-00256.x/abstract, 2001.

Smith, B., Samuelsson, P., Wramneby, A., and Rummukainen, M.: A model of the coupled dynamics of climate, vegetation and terrestrial ecosystem biogeochemistry for regional applications, Tellus A, 63, 87–106, https://doi.org/10.1111/j.1600-0870.2010.00477.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2010.00477.x, 2011.

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, https://doi.org/10.5194/bg-11-2027-2014, http://www.biogeosciences.net/11/2027/2014/, 2014.

Taylor, K. E., Williamson, D., and Zwiers, F.: The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, University of California, 2000.

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, https://doi.org/10.1046/j.1466-822X.2001.00175.x, http://onlinelibrary.wiley.com/doi/10.1046/j.1466-822X.2001.00175.x/abstract, 2001.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona-Moreno, C.: The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas emissions: results from a process-based model, Biogeosciences, 7, 1991–2011, https://doi.org/10.5194/bg-7-1991-2010, http://www.biogeosciences.net/7/1991/2010/, 2010.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R., and Schmullius, C.: Carbon stock and density of northern boreal and temperate forests, Global Ecology and Biogeography, 23, 297–310, https://doi.org/10.1111/geb.12125, http://onlinelibrary.wiley.com/doi/10.1111/geb.12125/abstract, 2014.

Tost, H.: Chemistry–climate interactions of aerosol nitrate from lightning, Atmos. Chem. Phys., 17, 1125–1142, https://doi.org/10.5194/acp-17-1125-2017, https://www.atmos-chem-phys.net/17/1125/2017/, 2017.

Tost, H., Jöckel, P., and Lelieveld, J.: Influence of different convection parameterisations in a GCM, Atmos. Chem. Phys., 6, 5475–5493, https://doi.org/10.5194/acp-6-5475-2006, https://www.atmos-chem-phys.net/6/5475/2006/, 2006.

Tost, H., Forrest, M., and Hickler, T.: Interactive vegetation influences on climatological meteorological fields and trace gas emissions, vol. 20, p. 12047, http://adsabs.harvard.edu/abs/2018EGUGA..2012047T, 2018.

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., Schwalm, C. R., Schaefer, K., Jacobson, A. R., Lu, C., Tian, H., Ricciuto, D. M., Cook, R. B., Mao, J., and Shi, X.: The North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project – Part 2: Environmental driver data, Geoscientific Model Development, 7, 2875–2893, https://doi.org/https://doi.org/10.5194/gmd-7-2875-2014, https://www.geosci-model-dev.net/7/2875/2014/, 2014.

Weiss, M., Miller, P. A., van den Hurk, B. J. J. M., van Noije, T., Ştefănescu, S., Haarsma, R., van Ulft, L. H., Hazeleger, W., Le Sager, P., Smith, B., and Schurgers, G.: Contribution of Dynamic Vegetation Phenology to Decadal Climate Predictability, Journal of Climate, 27, 8563–8577, https://doi.org/10.1175/JCLI-D-13-00684.1, https://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-13-00684.1, 2014.

Wramneby, A., Smith, B., and Samuelsson, P.: Hot spots of vegetation-climate feedbacks under future greenhouse forcing in Europe, Journal of Geophysical Research, 115, https://doi.org/10.1029/2010JD014307, http://doi.wiley.com/10.1029/2010JD014307, 2010.

Zaehle, S., Sitch, S., Smith, B., and Hatterman, F.: Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics, Global Biogeochemical Cycles, 19, GB3020, https://doi.org/10.1029/2004GB002395, http://onlinelibrary.wiley.com/doi/10.1029/2004GB002395/abstract, 2005.

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., and Norby, R. J.: Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies, New Phytologist, 202, 803–822, https://doi.org/10.1111/nph.12697, https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.12697, 2014.

Zhang, W., Jansson, C., Miller, P. A., Smith, B., and Samuelsson, P.: Biogeophysical feedbacks enhance the Arctic terrestrial carbon sink in regional Earth system dynamics, Biogeosciences, 11, 5503–5519, https://doi.org/10.5194/bg-11-5503-2014, https://www.biogeosciences.net/11/5503/2014/, 2014.