

Vegetation biogeography is a main source of uncertainty in modelling the land carbon cycle

Corresponding Author: Dr Ruiying Zhao

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Land carbon sink remain largely uncertain among intercomparison models, thus restricting our capability of predicting, mitigating and adapting to climate change. There have been research efforts which leveraged the emergent constraint approach to reduce the predicted uncertainty for the variables (i.e., photosynthesis or gross/net primary production - GPP/NPP and ecosystem respiration – ER). The essence of emergent constraint is to identify the emergent relationships (typically linear) between a multi-model ensemble and an observational estimate. Here this study by Dr. Zhao et al have used this emergent constraint approach to identify the remarkably strong linear relationship between carbon flux (GPP and ER) and biogeographic area of vegetation biomes. Then the two sets of remote sensing datasets (MODIS and ESA CCI PFT maps) were used as the observational estimate to constraint the uncertainty of GPP and ER, respectively. I am highly impressive by the approach of using biogeographic vegetation biome area and its strong relationships because this offers significant insights in term of vegetation biome impact on carbon cycling.

Here I have some moderate suggestions to improve this study. First, the key uncertainty or weakness (if any) is that the remote sensing data of vegetation biome area has large uncertainty as shown in the results of figure 1 et al. The authors have made a good job to clarify this. Since the emergent constraint largely depends on the accuracy of observational variable – vegetation biome area here. The authors may consider clarifying that 1) the emergent constraint could still reduce the uncertainty even if the two sets of remote sensing datasets were used together by combining the remote sensing datasets (i.e., by getting the range of minimum and maximum values of GPP and ER); 2) the needs for future studies to improve or have more reliable observational estimate of vegetation biome area. Second, the authors may also clarify the potential limitation of emergent constraint in term of underlying mechanisms underlying mechanisms (i.e., Winkler et al., 2019) and the spatial/temporal scales mismatch between data and models (i.e., Winkler et al., 2019). Ideally or necessarily, the research efforts in reducing the uncertainty of land carbon sink should be aimed by both improving the model structure or processes (i.e., the emergently vegetation biome distribution) and the emergent constraint approach here. This also related to the constraint of ER in this study because the current model structure have very simple structure for modeling ER which are strongly coupled with GPP. Third, the authors may also consider the potential effects of land usage impacts. Land use change likely have influenced the vegetation biome distribution more than climate change in term of spatial and temporal scales. Currently, it is not clear how the land usage change has been accounted for.

Since the paper is very well written with insightful discussion, I also have the following minor suggestion for the authors to consider

Abstract: may also consider reporting the results of ER

Lines 44-45: it would be important to mention competition, i.e., by accounting for size or cohort based competition for resources (light, nutrients and water)

Introduction: may consider clarifying what the emergent constraint is and the limitation of emergent constraint.

Figure 1: the color scheme would need to be improved. Also there are different definitions for vegetation biomes such as WWF vegetation biome. My knowledge is also that there is 1 km resolution of land cover maps

<https://www.earthenv.org/landcover> I am not sure whether this is useful as an additional observational constraint.

Overall, the whole manuscript has focused on the constraint of GPP, but the results of ER was indeed also important for carbon sink. Thus, the authors may consider balancing this.

Figure 2: I am curious whether N deposition has been accounted for and how important it would be.

Discussion: as clarified, the authors may consider clarifying the limitation of emergent constraint and needs of improving model structure/processes for the mechanistically deriving/model the emergent properties of vegetation biome. The section of discussion on structure (vegetation biome biogeographical distribution) vs biogeochemistry is interesting. It is also good to learn that the authors used emulator to disentangle their relative impacts. With this, the authors may consider elaborating a bit more for the implication of improving the model mechanisms in term of more research efforts on vegetation biome biogeographical distribution itself vs biogeochemistry.

References

Winkler, A. J., Myneni, R. B. & Brovkin, V. Investigating the applicability of emergent constraints. *Earth Syst. Dyn.* 10, 501–523 (2019).

Reviewer #2

(Remarks to the Author)

Reviewer's comments

This paper attempts to identify an observational constraint for dynamic global vegetation models (DGVMs) by establishing a linear regression relationship between plant function type (PFT) area and total gross primary productivity (GPP) for a given PFT. The authors use satellite-based observations to infer a plausible range of total global GPP per-PFT (GPP_{pft}). They correctly identified that land surface models with dynamic vegetation components are often deficient when it comes to simulating present-day biogeography, and that using these PFT observations to constrain total global GPP estimates can tighten the range of GPP uncertainty.

I have serious concerns, however that the linear regression on which this study is based is fundamentally flawed because both variables used in the regression have a strong dependence on PFT area. While I broadly agree with the conclusion that observed global area of a PFT can be used to refine global GPP estimates derived from DGVMs, I have serious reservations about the approach taken here. My criticism boils down to the fact that in this regression relationship, the independent variable (x-axis; PFT area) is also a central component of the dependent variable (y-axis; total global GPP_{pft}), and therefore the resulting strong relationships shown for each PFT are not valid because it is essentially using PFT area to predict a variable that is strongly dependent on PFT area. In fact, the only reason why there is variation in the relationship (and not a perfect 1:1 line) is because there is variation in the per-model and per-grid cell GPP_{pft} rate per unit area. In order to derive total global GPP_{pft}, I assume the authors would have had to multiply the raw land surface model output GPP_{pft} rate per unit area (likely in units similar to gC m⁻² yr⁻¹) by the grid cell PFT area, and then sum all grid cells (possibly also making a multi-year mean of the timeseries). Similarly, converting model output of PFT fraction to a total land area for a given PFT (Areap_{pft}) would involve multiplying the per grid cell fraction by each grid cell's area (giving the exact value used in the GPP calculation in the previous sentence), and then summing all grid cells. If you take out the area dependence from the simple GPP calculation, you are left with the GPP_{pft} rate again per unit area. Therefore, I think the linear regression on which the emergent constraint is based is flawed.

Having said that, I do think there is value in highlighting that observations of PFT area can be used to constrain the simulated GPP of models in order to gain a more accurate measure of present-day GPP and consequently atmosphere-land carbon fluxes. I just don't think that it can be done in the way that the authors have proposed. I believe that the authors would learn more about the models and the plausible constraints on GPP by looking in detail at the locations where PFT fractions match observations, and those where PFT fractions do not. I would also encourage the authors to examine other carbon-cycle emergent constraint studies and consider the definition presented in the Cox et al. (2019) paper:

"The emergent constraint technique involves using the full ensemble of models to find an across-ensemble relationship between an observable feature of the Earth System (such as a trend, interannual variation or change in seasonality) and an uncertain aspect of the future."

I would argue that while PFT fraction is an observable feature of the Earth System (and GPP the uncertain aspect of the future), I don't think that it provides a trend, interannual variation or change in seasonality, in the same way that other emergent constraint studies have. See for example, Cox et al. (2013) who found a strong linear relationship between the sensitivity of tropical land carbon storage to warming and the sensitivity of the annual growth rate of atmospheric CO₂ to tropical temperature anomalies.

Other Comments

Line 1: Title. I'd suggest reviewing the title, as I think "main source of uncertainty" implies that you've compared biogeography to other sources of uncertainty in the land carbon cycle.

Line 25: Worth making clear that you've reduced the spread of land carbon estimates for the present day, or does it also cover the future?

Line 26: Again, emphasise that you're talking about historical changes

Line 33: the term "offsetting" tends to be used in carbon trading. I'd suggest being more quantitative here – perhaps cite the fraction of anthropogenic CO₂ that the land absorbs?

Line 54: Limitations could also include consistency of satellite timeseries, particularly as we're interested in the response of the C cycle to increasing CO₂ emissions

Line 58-59: See line 1 comment.

Line 76: Maybe clearer to say "global" GPP here

Line 82: Would perhaps be better to assess how improved estimates of GPP affect the uncertainty in land carbon fluxes, and improve our ability to close the C budget over this period

Line 87: Is the PFT total area also in km² / year?

Line 90-92: The premise of the whole paper hinges on this statement – see main reviewer comments.

Line 93: Why didn't the authors split EBF into tropical and temperate in a similar way to grasses? There would likely be a huge range in the GPP rate within the EBF PFT.

Line 93-101: How do these values relate to the annual rate of GPP that would have come directly out of the models? Within a given PFT distribution, there would be a range of GPP rates (measured in similar units such as gC/year/sqm), so how are these rates different from just using the observed PFT distribution maps and taking a spatial mean and standard deviation of the modelled GPP rate?

Line 101-103: This statement is probably for the discussion, and needs referencing / supporting with evidence from study

Line 177: This relationship is bound to be strong if both variables co-vary strongly with area

Line 229: Emphasise that it is PFT-specific total GPP

Line 280-281: I don't think you can conclude this from your results. I would argue that the higher uncertainty reduction in the approach presented in figs 1 & 2 is due to the over-dependence on area in the total GPP vs PFT area relationship

Line 282-283: Again, I don't think you can conclude this from your results. There is a lot of literature that shows parameter tuning from detailed site-specific observations - see for example Harper et al who used plant trait information to tune JULES, in places where the PFT fractions were very well documented. <https://gmd.copernicus.org/articles/9/2415/2016/gmd-9-2415-2016.pdf>

Line 331: I think it's important to be specific here - your study attempts to understand and reduce uncertainty in the *historical* to present-day carbon cycle. It highlights that DGVMs have work to do to improve the modelled spatial distribution of the their PFTs.

Line 350-351: Worth acknowledging (apologies if I've missed it) that in using the S2 scenario, we would expect DGVMs to over-estimate the fraction of tropical broadleaf evergreen tree (which is the PFT class that normally has the highest contribution to global GPP)

Line 359: See earlier comment on Tropical vs Temperate EBF

Line 378-379: Maybe a better description of the metric is need here, given the >1200C criteria, I would assume it is rather something like accumulated temperature during the growing season?

Line 417-418: Is this relationship on an annual timestep?

Line 419: I think I'm right in saying that β_i is the GPP rate per unit area here. In that case you would also need a sigma to sum across grid cells.

Reviewer #3

(Remarks to the Author)

Zhao et al present a method for constraining total global GPP and also to explore relationships between change in GPP over time with changes in PFT area. To constrain global GPP the authors define linear relationships across models between total global PFT specific GPP and PFT area and then constrain the range in global GPP values based on independent (though see caveat on that below) datasets of PFT area. In so doing, the authors highlight the importance of PFT area in modeled global GPP uncertainties. They also (for one independent dataset) constrain the mean global GPP to a higher value than is often reported, but which is more in line with other independent total GPP estimates.

I served as a reviewer of the original version of this manuscript. I appreciate the extensive additional work the authors have done to revise the manuscript and to add extra, very involved analyses to address some of the concerns.

In my original review I emphasized that the goal of this study, to highlight the importance of differences in PFT area on uncertainties modeled GPP estimates, was much needed. I still think that, and furthermore I think the changes the authors have made to the paper have enhanced that message. I am very supportive of seeing this paper published.

In the original review I had concerns about the approach that have mostly been addressed. The difference in framing alone helps me to better digest and have greater confidence in the results and conclusions of this study. However, I still see some caveats that affect the strength of the conclusions drawn. I highlight these below, followed by key suggestions to address what I see as the remaining caveats. Following this general comments section I provide additional suggestions for several of the methods and results section before finishing with some minor comments.

Please forgive any repetition in the sections below. I have reviewed this paper over the course of the week because like this paper and I think it's important and therefore I wanted to provide a thorough review. This required me thinking through some parts of the new analyses and responses to reviewers and took longer than I expected. I have tried to tidy up my comments but in a rush to get this review to you I may have missed some repetition. Apologies for that. I would be happy to discuss any of these comments further, either via the journal editors or directly via email.

-- Natasha MacBean (nmacbean@uwo.ca)

Remaining caveats with the approach used and therefore strength of the conclusions drawn:

- I think the understanding gained from the linear relationships between total GPP and PFT area are more nuanced than the authors lay out in their paper (more details below).

- I think the uncertainties on the PFT maps are too narrow and therefore the large reduction in total GPP spread obtained by using the PFT area from these two maps is not realistic (more details in methods and results below).
- The fact that the two different LC maps result in widely different mean total GPP estimates, even with (what I think are) too narrow uncertainty bounds, highlights the issue that the reduction in spread of total GPP values from constraining with satellite derived PFT area (as shown in Figure 2) is unrealistically high. Therefore the strength put on this statement – the reduction of 86% – just does not seem to be reliable.
- There is a circularity in using the ESA CCI map at least, because some models are using that map to prescribe PFTs.

Suggestions to address these caveats:

- I will let the authors digest what I have said about the R2 values (especially those not constrained to go through zero), the strength of the relationship between total GPP and PFT area, and what (I think) that says about our understanding of differences in PFT representation across models. But I think that there needs to be a more nuanced description and discussion of those results (and this section is currently very short). I would not place as much strength in the statement in lines 90-92 that “The strong relationships indicate that the total GPP of each PFT across DGVMs is largely dependent on its prescribed or simulated area, and that DGVMs simulate a similar per-area GPP for each PFT”.

- I would like to see more realistic, wider uncertainties for each of the PFT products that included LC mapping uncertainties propagated through cross-walking uncertainties to the PFT maps (unless I'm missing this here – see more detailed comments in methods below), and therefore wider ranges of global total GPP (and a lower reduction on the total GPP spread) as a result. However, I would not wish that task of propagating cross-walking uncertainties to PFT maps on anyone! It was difficult enough in the Hartley et al. (2017) paper and we only explored some of the possible uncertainty “pathways” to demonstrate the issue – not the full range of uncertainties (which again leads me to think cross-walking uncertainties has not been accounted for here). So at the very least, I would like to see a clear statement and discussion that these error bars in Fig. 1 do not incorporate all sources of uncertainty in the final PFT maps, especially not in relation to the highly subjective cross-walking uncertainties, bioclimatic limits and somewhat arbitrary phenology thresholds used to link between LC class and PFT and to differentiate PFTs of different biomes, and therefore the real range could be much higher. Still, and this is a key point, even acknowledging wider uncertainties on the PFT maps, this could still constrain GPP to a higher value than is currently assumed (120-130PgC) – for MODIS at least.

- Further to the previous point, I would like to see more discussion of all the sources of uncertainties in satellite derived PFT maps and what we as a community could do about them. This is mentioned in the discussion, but if there is space (I guess not) I would like to see more. Simply because the authors clearly do show that PFT area is an important source of uncertainty – that much is without a doubt for me. So which product is the best to use to prescribe PFTs in models, or to evaluate dynamic vegetation models (or to use as a constraint in the approach used in this study)? Do we just need an extensive worldwide campaign of validation of PFT maps – perhaps using citizen science? What else can be done? Examples of using ancillary datasets of tree cover fraction, as has been done with ESA CCI, sound good to me, but I am sure there could be other approaches. Perhaps we could explore the idea of optical plant functional types (but see minor comment below in response to R1C3). Also we could use the approach you have taken with emulators to fully explore more extensive PFT map uncertainties.

Methods – DGVMs

It's not clear from the text, unless I missed it, that some of these models have prescribed, static PFT fractions and some have dynamic vegetation. It reads in the methods text as if all the models have dynamic vegetation, but that's not the case is it? In fact, from reading the first section of the results it becomes clear that some models have prescribed and some simulate their dynamic vegetation. In this format of paper people may read the methods after the results, but for clarity I would add the information about whether the models have prescribed or simulated vegetation fractions to the DGVM methods section (and/or maybe Table S1)?

Following on from this point – if most models have prescribed vegetation then the statement that parameters feedback on PFT distributions is not correct for those models that do not have dynamic vegetation enabled, and this confused the additional results that are presented in Figure S5. More on this below.

Methods – emulators

I really like the approach of developing PFT specific emulators to test PFT uncertainties. I appreciate this is an additional analysis prompted by a suggestion in my original review. I think it's a very novel approach, so much so that there's part of me that thinks it's such a nice and novel analysis that it could be its own paper.

However, I need more information here. What were the exact inputs to the emulator? Over how many grid cells was each PFT specific emulator developed (e.g. grid cells selected across a range of climate zones for each PFT?) and if so, how were those grid cells selected? Or was it a global emulator for each PFT? What GPP data were used to train the model? Over what time period were the outputs of the random forest model? Was a spin up and transient included in the emulator? Was each emulator independently tested/validated? If so, what were the results of that? All of this feeds into how we can interpret Fig. S5.

Methods – Remote sensing estimates and uncertainties of global PFT distribution

I appreciate the consideration the authors have taken in improving the methods section by adding details and clarifications. I am still a little uncertain how the uncertainties in both the underlying LC map, and the “cross-walking” to PFTs (including uncertainties in bioclimatic limits and phenology based thresholds to separate out different climate zones) has been considered. The authors discuss this, but a) I still don’t understand some details; and b) I still don’t think all these uncertainties have been considered.

For example:

“We first obtained the uncertainties generated during mapping each land cover type by machine learning classifiers, and then assigned the uncertainties to PFTs based on translation rules from land cover to PFT maps (Bonan et al., 2002).” What are the machine learning classifiers in this case? How does that give you the full uncertainty related to the LC map? And have uncertainties related to the translation rules between land cover and PFT map in the MODIS product also been considered – I can’t see any consideration of uncertainties in the cross-walking from LC to PFT in the Sulla-Menashe? In fact at Lines 407-408 the authors say “The uncertainty in MODIS PFTs was based on the reported uncertainties in the initial land cover mapping (Sulla-Menashe et al.)”. This explicitly states that only uncertainties in the LC maps have been considered. Not any uncertainties in translating those (often broad) LC classes into PFT fractions. As we described in Hartley et al. (2017) the range of % you can give for some classes, especially in the very mixed and sparsely vegetated cover types, is almost impossible to know and very subjective. Similarly, you used additional bioclimatic limits from Bonan et al. (2002) and phenology based thresholds to further define PFTs according to their climate zone. However, these bioclimatic limits and phenology based (e.g. GDD) thresholds have uncertainties as well, and this is also not represented in your MODIS error bars. Some of the phenology based metrics – e.g., whether you could GDD above a temperature threshold of 0 or 5°C – are very arbitrary.

Also:

“Uncertainty for the ESA CCI product was derived by digitizing reported uncertainties” what does digitizing reported uncertainties mean? So I see how the ancillary tree cover dataset can be used to estimate tree %, but what about all the other LC classes such as sparse vegetation – how did you assign fractions to each PFT? Did you use the ESA CCI cross-walking table for that? And therefore did you consider uncertainties in those very subjective cross-walking values, as described in Hartley et al., 2017.

Results – Relationship between PFT GPP and area

I understand perhaps better than I did in the original review what point the authors are trying to make here. Still, I would say that for some PFTs the linear relationship is not as strong (lower R² values) and the fact that the R² values are often lower when they are not forcing the regression line through zero also gives me pause. The logic for forcing the regression through zero is sound, but the authors are also just trying to see what the sensitivity of the model GPP is to the PFT area. In cases where the R² is much lower, then actually what the non zero constrained regression is telling you is that actually the total GPP is not that sensitive to the area of the PFT as it is for other PFTs. This is especially the case for C3 grasses which are one of the most abundant PFTs worldwide (if you add up the area of the individual C3 categories in this study). So when the authors say in their response to R1C2 that “it is very surprising to see ALL models show EBF has similar productivity per area (i.e., the slopes of the total GPP – area relationships for each PFT in Fig. 1)” I would say I agree, but I also think that’s not the case for all PFTs.

I think this analysis is interesting though because the differences across PFTs in terms of the strengths of the linear regression (unconstrained R² values) provide insights into which PFTs have very similar representation across different models (and therefore where the prescribed PFT area is causing most of the difference in total area) versus the PFTs where area plays a role, but clearly other aspects of modeling that PFT are contributing to the spread in total GPP (which then gives lower R² values).

Some of the PFTs that clearly have a strong linear relationship between total GPP and area are not that surprising – i.e. C4 grasses have had little attention paid to them compared to C3 grasses, and so it is perhaps predictable that all models simulate a similar GPP per area as their underlying representation is very similar. In contrast, more work has been done on C3 photosynthesis – for example, different models for calculating C3 photosynthesis, different parameter values (some of which have been calibrated against more numerous datasets for C3 grasses), including optimality-based approaches, and including temperature acclimation, etc – and therefore I assume there are differences across models, and that explains the lower R² values. (I also note that Figure S1 is not cited in the main text). Deciduous needleleaved forests are another example that have had relatively little attention paid to them compared to other PFTs and therefore the R² values are quite high.

Results – Updated land carbon cycle estimation using remotely sensed PFT area

Given the issues I mentioned above I am afraid I do not trust the narrow range of uncertainties used for either MODIS or ESA CCI. See my comments in the general section on suggestions for addressing the key caveats above.

Fig. 2: Caption include the fact that the DGVMs are from TRENDY and should say which version, even though this is in the methods (as many people won’t read the methods). In fact I would make it clear what is going into each of these estimates in

the figure caption. Which products are going into MLEC for example, just FLUXCOM or a range of different upscaled EC products?

Results – Contribution of PFT change to global GPP change

This section is really nice and interesting and yet it is a little lost in the overall message about constrained PFT area leading to constrained GPP based on the strength of the linear relationship between PFT GPP and area. I would have liked to see more here, although I appreciate there's not much space in a letters journal. What is very curious about this is that we expect to see C4 grass outcompete at higher CO₂ concentrations, if I'm remembering correctly. This is not directly discussed although the authors mention the potential water savings under elevated CO₂. I would suggest the authors expand this though to suggest future studies check that that is indeed what is happening (water savings) and the increased C3 is not because of incorrect process representation in the models.

Discussion

Lines 247-249: "To improve biogeography process in DGVMs, we suggest prioritizing efforts to clear hotspots of biogeography uncertainty (e.g., those identified in Fig. 3) and to develop a more reliable global PFT map" Yes absolutely! Actually most of these models have no dynamic vegetation nor vegetation demography that is causing any kind of PFT change. They all have prescribed maps. So yes, a more reliable global PFT map is needed for models that do that! And yet, as the authors have shown, the two most widely available maps have very different areas, which, as the authors have also shown, feeds into differences in total GPP (though I just wouldn't place as much emphasis on the narrowing of uncertainty in that total GPP estimate as the authors do). This is the key message of the paper here I think, but I believe it can be framed more clearly.

Line 249-250: I guess I would say be careful with the use of "alternative" in this statement: "Using a consistent biogeography baseline, such as a PFT map from remote sensing as a diagnostic alternative" because these PFT maps are being used in models. I know ORCHIDEE and SDGVM use the same ESA CCI map (which also makes me wonder about the circularity in this approach – I would like the authors to comment on that).

Lines 256-261: This is only true for dynamic vegetation models, of which there are only 4 in the TRENDY v9 ensemble. Otherwise, photosynthesis rates etc do not feedback on PFT area. So I would revise this section accordingly.

Lines 269- : I like the additional emulator analysis and I really appreciate the authors attempting to address this comment from the first round of review. I have one minor comment and one concern:

My concern is that I don't think the larger spread (smaller reduction in uncertainty in EM(DGVMs,MODIS)) is because the models have tuned their structure to the wrong location. Or at least, I can't see how that would be the case. Let's say C4 grass is in the completely wrong location in many models because the bioclimatic limit used to separate out C3 and C4 grasses is wrong. If so, what does that imply, that their V_{cmax} parameter for example is adjusted to the climate of the wrong location to match in situ flux tower data? That seems convoluted to me, although I think maybe I see the argument. I'd be willing to bet though that most PFTs are not in the completely wrong location (mostly based on prescribed PFT maps). It would mean major biases in parameters or model structure to account for that. What I think is maybe more likely, or explains more of the difference between EM(DGVMs,MODIS) and DGVM_MOD is what I've said throughout this review – that the uncertainty reduction from constraining with MODIS (or ESA CCI) PFT map area is too high. I would guess that if you had more realistic uncertainties on your PFT maps, your constrained GPP would also be about 40-50% as in EM(DGVMs,MODIS). It still shows a big importance of PFT map, but again, I think it just cannot explain the majority of the spread across models. Still, this nice emulator result shows that differences simply due to the PFT map are a greater issue than anyone would bet.

My minor comment is that I am not sure I would describe it as a difference between biogeography and biogeochemistry (Responses to R1C7 and R2C3 and new text describing Fig. S5) (with my issue being mostly with the use of biogeochemistry). If the location of the PFTs is wrong (e.g., Bias type 1 in Response to R2C3), then that's also a biogeography issue right? Or am I misunderstanding? If the location is wrong then it could be that the bioclimatic envelopes for example are not right, and I would call that a biogeography "process" or limitation (that is intimately tied to area). On the other hand, you may have an issue with temperature optima, which would prohibit certain PFTs from growing in an area. I would call that a biogeochemistry (or plant physiology actually) process (could also be related to phenology thresholds or ecohydrology if not enough water was available). So it's just a question of wording but I perhaps would more simply say that what you said in response to R2C3 – that there are two types of biases related to area and location, and that biases related to area only were accounted for by the analysis in Figure 1. But Figure 3 showed the second bias related to location distribution, and so you have done this analysis. Basically, I think the use of biogeochemistry here is confusing and also misleading given that for most models any "biogeochemistry/plant physiology" etc process is not feeding back on the distribution anyway. If the distribution is wrong for most models then it's because the prescribed PFT map is wrong (except for the dynamic vegetation models).

Reply to response to R1C3 in response to reviewers: I would perhaps disagree that PFTs can be directly linked to spectral signatures. In fact, it is very hard to distinguish trees from grasses, even in hyperspectral data and even when you add LiDAR data (as my PhD student Rubaya Pervin found in drylands given there's also a strong spectral signature from dry grasses and/or biocrusts). In the past this has prompted the suggestion of using "optical functional types" (Ustin and Gamon, 2010) where global vegetation would be classified based on the clustering of spectral signatures. To my knowledge, no one

has really tried this at regional or global scale. I would bet this would look different to the PFT maps we currently use and would be hard to match with anything we represent in models. So the question remains, how can we use satellite data to infer plant functional type fractional cover? The fusion of tree cover fraction in the ESA CCI product is a great advance. But it also depends on what is counted as a tree (height thresholds etc).

Reply to response to R1C6 in response to reviewers: Yes this is very interesting to see the range of estimates, and also the reporting of more recent studies that all have higher estimates. In fact, the higher mean value of GPP as constrained with MODIS is one of the most interesting things about this paper (although this is somewhat dampened by the fact that ESA CCI maintains the typical value of around 120PgC), although other methods have a higher value as the authors show in Fig. 2a. As I am signing this review I apologise for this shameless self-citation, but a paper in which I constrained GPP using SIF data (one of the earlier, simpler approach using just one model) also found a posterior (post parameter optimization) global value of just over 160PgC (MacBean et al., 2018). Please do not feel you need to include this as a reference, it is more for your interest. Since I wrote that paper in the back of my mind has been the thought that yes ok we constrained the high positive bias in ORCHIDEE (pre optimization it was above 190PgC), but we have not constrained it enough. I hadn't quite processed until I read the authors response to the original review that there are indeed multiple studies pushing at this higher value (and hadn't seen the Liu paper). As the authors say, the value of 120-130PgC has been talked about for Beer's 2010 Science paper if not before and is the number that is most often quoted. It has been 130PgC in the the GCB for at least the last two years. So I think it's great that the authors have moved away from a "correct" magnitude and instead show the uncertainties. Still, when you put all these studies together with the new study presented here, it does feel like the evidence is mounting for a higher value.

Minor points:

Line 29: would suggest "biogeography-climate" instead of "biogeography – climate" interactions. First time I read it I thought the words after the hyphen were making a different point.

Lines 54-56: another limitation for MODIS GPP (or any GPP derived from the production efficiency/LUE method) and the machine learning approaches I believe is the lack of accounting for the effect of photosynthesis on LUE (De Kauwe et al., 2016).

Line 71: would maybe not use the word "prescribe" here (to be pedantic) – maybe just that in process-based models these processes are specific to each PFT ?

I cannot see CLM5 in Fig. 1j? I'm guessing that's why you have axes upper limits that are much higher than the main cluster of models

I like the new additional analysis presented in Fig. S6. Nice!

Data Availability Statement needs to include where you got the TRENDY v9 GPP and PFT from as it's actually not on the global carbon budget page that links to Mike O'Sullivan's Github (<https://mdosullivan.github.io/GCB/>). Also, would you consider providing the emulators in a publicly available repository for people to test other PFT maps?

References

- De Kauwe, M. G., Keenan, T. F., Medlyn, B. E., Prentice, I. C. and Terrer, C. (2016) Satellite based estimates underestimate the effect of CO₂ fertilisation on net primary productivity. *Nature Climate Change*, 6, 892-893.
- MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., ... & Disney, M. (2018). Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. *Scientific reports*, 8(1), 1973.
- Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, 186(4), 795-816.

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I am highly impressed by the revision. The revised manuscript has substantially addressed my concerns in terms of uncertainty of vegetation biome areas by using more data products and also potential impacts of land use change. The writing has been also substantially improved. Here I only have one minor suggestion for Dr. Zhao et al to consider. As commented in the first round of review, it would be useful to clarify some caution (i.e., by needs of incorporating underlying mechanism in future studies) of this data-based approach to reduce uncertainty of projected carbon cycling - GPP and ER here. Vegetation biogeography is emergent property in which plant individuals grow up and compete for the resources. With this, it could have uncertainty based on the transient snapshot of relationship between vegetation biome area and GPP & ER at community scales for projecting future GPP and ER dynamics.

Reviewer #3

(Remarks to the Author)

I am satisfied the authors have addressed most of my concerns/comments and I thank them for taking the time for that. I especially appreciate their consideration of my point about RS product uncertainties and the fact they now use the spread across 5 RS products. However, I still have concerns about the use of a simple linear regression to highlight the main point of this study, which I think mirrors some of Reviewer 2's concerns, though I don't share Reviewer 2's concern that it's wrong to have GPP as the dependent variable and PFT area as the independent variable.

I carefully considered the comment from Reviewer 2 about the scaling of GPP with area. I do not see their problem with having total GPP dependent on PFT area, statistically speaking, given GPP is the dependent variable and the PFT area the independent variable, unless I'm missing something here. However, having considered this more, I do perhaps see some of Reviewer 2's issue with using the linear regression in this way. I've tried to articulate some of my concerns below.

First, I think one thing that Reviewer 2 is pointing out is that it is obvious that GPP scales with area. I appreciate (and agree with) the authors' response that although it seems obvious given the equation for global total GPP, PFT area is not that discussed as a source of uncertainty in global GPP in the modeling community, and instead studies focus on differences in the way the model represents the photosynthesis, phenology etc (biogeochemistry as the authors discuss). But unfortunately I still think some readers will say it's obvious given the dependence on area. Further, let's say some models have extremely different area, so the range in PFT area is high, then the linear relationship will become more pronounced despite variance around the regression line due to differences in model structure/parameters etc (biogeochemistry) in estimating GPP per unit area.

I was somewhat underwhelmed with the author's response to R3C6, and again I think this speaks to my, and perhaps Reviewer 2's, sense that the emphasis placed on the results in this study are more nuanced than the strong statements in lines ~92-96 based on a simple linear regression. In R3C6 in my last review I questioned forcing the linear regressions through zero, and the fact that for some PFTs the R^2 values are much lower if you don't force through zero. This wasn't addressed in the authors' reply to my comment in the last round of review. Again, while it is true in reality that with zero PFT area then GPP will be zero, here the authors are trying to understand model behavior, and therefore I'm not sure these linear relationships should be forced through zero. It's hard to tell for sure, but in most cases where the R^2 values drop when the regression is not forced through zero it looks like the slope of the regression would be lower. That would actually give you a narrower range for the GPP values (higher reduction in GPP range), but it's hard to know where that slope would lie (what the y-intercept would be) and therefore what the actual GPP range would be. It could be quite different I think. So for full transparency I think the authors should not only report the R^2 values of the regressions not forced through zero but also the actual regression line (with slope and y-intercept) and therefore how the GPP range would be different with a regression line not forced through zero. And then see if the key points of this paper – the reduction in uncertainty on GPP (and ER) and the constrained range of GPP values – changes as a result of not forcing through zero.

In response to my point in R3C6 about the fact that some PFTs do have lower R^2 values (and therefore again the strong statements in lines 92-96 seem a little overstated) the authors simply added "the relatively lower R^2 values, such as those for C3 grasses and crops, may indicate greater inter-model difference in LAI or leaf photosynthetic rates compared to other PFTs." First, I wouldn't use "may" here. If there's spread around that line it must be that there's an inter-model difference in how the models are simulating other components of total GPP such as phenology and photosynthesis, plant water uptake, nutrient dynamics etc (factors that contribute to A_{leaf} and LAI). But second, I'd appreciate seeing a more nuanced discussion about the fact that lower R^2 values for some PFTs mean for those PFTs there is more disagreement among models in the representation of the biogeochemistry/physiology etc. The point is that where we don't have that much agreement among models, PFT area will become a much stronger predictor of total GPP because the models don't differ that much in their parameterization of the biogeochemistry (perhaps for reasons I gave in the last round of revision, such as PFTs that haven't had as much focus on how we model them, such as C4 grasses). Again, a reader may think this is obvious and so I urge the authors to discuss this in a more nuanced way.

Finally, why didn't the authors do a multivariate regression with all three components and report semi-partial correlation coefficients? I apologise I'm only asking this now in a 2nd round of review, but the comments from Reviewer 2 have prompted me to think more on this. They do show the R^2 of the simple regressions between each pair of variables and I don't think overall the rank order of what is important will change. I also appreciate that PFT area is likely not correlated much with A_{leaf} or LAI (though those two components are likely correlated), especially for models with prescribed PFTs so a simple linear regression suffices. But for models with dynamic vegetation PFT area could be correlated with A_{leaf} and LAI and so knowing the semi-partial correlation coefficients would be interesting. I think this also might help satisfy the readers who might be tempted to think the single variable linear regression is too simple? I think the multivariate regression

would also take care of the problem of forcing the regression through zero in the PFT area vs GPP case.

I also like Reviewer 2's suggestion about looking at the spread in GPP values across for locations where modeled PFTs correspond well to each other (e.g. low CV in Fig. S5) and the spread for locations where modeled PFTs do not correspond well to each other (higher CV in Fig. S5). I hope the authors take these concerns and suggestions in the spirit they are intended. As the authors note, very few people consider that PFT area is an important component of spread in total model GPP. But I fear that with the simple linear regression, readers may be quick to try to dismiss the claim that PFT area is indeed important.

I am more persuaded by the authors' argument related to the emulators that if you "correct" the PFT location (not just the area) and then get a lower reduction in uncertainty then you can postulate that is because the PFT parameters (or sensitivity to climate of that location) have been tuned to the wrong location. However, I'm not sure I'd phrase it as PFT locations were "correct" just because the authors have used the same prescribed PFT for all model emulators. Perhaps "consistent" would be a better word? Also, MODIS is not necessarily correct, and therefore couldn't it be the case that MODIS is incorrect, so now the models do not have the correct PFT in that location? Thus having a more incorrect PFT location by using MODIS could be contributing to the lack of reduction in uncertainty, unless I'm misunderstanding something? I like Fig. S6 (although note that it's not cited in the text), and I appreciate the extra analysis and response to this point, but there are not many models that have PFTs in regions we're not expecting. Looking at Fig. S5 the highest CV across models tends to occur at the boundaries of the typical ranges of each PFT, suggesting to me that it's more about slight shifts in distributions than completely wrong locations/different climate? What I will say is that this is nonetheless an interesting discussion point. If what the authors suppose here is true, then it has implications for how we optimise parameters – it suggests we should not be using RS maps to set PFT fractional cover in parameter optimizations. Even if RS data are used, the PFT fractional cover for that grid cells should be independently validated with field data. This in turn has implications for using complete gridded satellite data (for example all grid cells worldwide) to optimize parameters. This can be easily tested with two optimisations using differing PFT fractional cover and I'm not sure this has really been done before (not suggesting the authors do this at all – merely thinking on paper).

One more question about the emulators, just for clarification. Were the emulators were run for each grid cell and the pft fraction for that grid cell was mapped to the pft dependent GPP for that grid cell? Or was it scaled up so it's total GPP corresponding to total PFT area?

There really seems to be an issue with crop cover in the models, if I'm not misunderstanding? And a wide range across RS products for C3 crops. I'm surprised by this. Again, more thinking on paper.

Line 204: did you mean to cite Fig. 3 here?

made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Reviewers

Reviewer #1 (Remarks to the Author):

R1C1: Land carbon sink remain largely uncertain among intercomparison models, thus restricting our capability of predicting, mitigating and adapting to climate change. There have been research efforts which leveraged the emergent constraint approach to reduce the predicted uncertainty for the variables (i.e., photosynthesis or gross/net primary production - GPP/NPP and ecosystem respiration – ER). The essence of emergent constraint is to identify the emergent relationships (typically linear) between a multi-model ensemble and an observational estimate. Here this study by Dr. Zhao et al have used this emergent constraint approach to identify the remarkably strong linear relationship between carbon flux (GPP and ER) and biogeographic area of vegetation biomes. Then the two sets of remote sensing datasets (MODIS and ESA CCI PFT maps) were used as the observational estimate to constraint the uncertainty of GPP and ER, respectively. I am highly impressive by the approach of using biogeographic vegetation biome area and its strong relationships because this offers significant insights in term of vegetation biome impact on carbon cycling.

Response: We thank the reviewer for their continued support of our study and for providing constructive feedback. We have endeavoured to use their comments to further improve our study.

R1C2: Here I have some moderate suggestions to improve this study. First, the key uncertainty or weakness (if any) is that the remote sensing data of vegetation biome area has large uncertainty as shown in the results of figure 1 et al. The authors have made a good job to clarify this. Since the emergent constraint largely depends on the accuracy of observational variable – vegetation biome area here. The authors may consider clarifying that 1) the emergent constraint could still reduce the uncertainty even if the two sets of remote sensing datasets were used together by combining the remote sensing datasets (i.e., by getting the range of minimum and maximum values of GPP and ER); 2) the needs for future studies to improve or have more reliable observational estimate of vegetation biome area.

Response: Thank you for your suggestions. Please kindly see our detailed response below:

1) We agree that the accuracy of remote sensing (RS) plant functional type (PFT) maps could affect the resulting GPP range. Though in the last round we have clarified that the uncertainty in RS PFT would not affect our main message as long as the strong GPP-area relationship stands, we now push further on this issue at the request of the reviewer. In this round, we collected five widely-used RS PFT maps (Table S3), including those from MODIS and ESA CCI.

We have now updated our main results (Fig. 1, 2, 3) using the ensemble of five RS PFT maps to constrain GPP instead of just using MODIS and ESA CCI PFT maps. After considering the inter-product difference in RS PFT maps, we found the GPP estimate constrained from 146.32 ± 26.39 PgC/year to 135.66 ± 6.54 PgC/year and ER from 143.39 ± 25.99 PgC/year to 133.01 ± 6.53 PgC/year, a 75% reduction for uncertainties in global GPP and ER estimates. This reduction is less than when only using the MODIS PFT map (86%), but is still very substantial and supports our main conclusion that vegetation biogeography is a main source of land carbon cycle modelling.

2) We fully agree that it is critical for future studies to improve global PFT mapping, in particular future studies need to ensure the consistency across products to better support the land carbon cycle modelling. Using the newly collected five RS PFT maps, we identified the hotspots of biogeography disagreements among RS PFT maps (Fig. 3). We suggest that future study should prioritize efforts to clear these hotspots of PFT disagreements (L212-220) and our additional analysis also highlighted the PFTs that we need to work on, such as C₃ cropland and shrubland (Table S3).

Table S3. The PFT area estimates across five remote sensing-based products (unit: 10^6 km²). Only MODIS and ESA CCI PFT maps have reported part of their own uncertainty. σ refers to inter-product uncertainty (i.e., one standard deviation of the area from five products).

PFTs	MODIS	unc	ESA CCI	unc	CGLS_ LC100	unc	GLC_ FCS30D	unc	GLASS- GLC	unc	σ
EBF	21.56	1.23	14.23	2.85	18.30	NA	15.30	NA	13.97	NA	3.23
ENFt	5.08	1.21	4.17	0.83	7.32	NA	5.24	NA	2.61	NA	1.72
ENFb	4.69	1.11	3.41	0.68	6.20	NA	5.13	NA	8.17	NA	1.79
DBF	16.13	2.79	7.21	1.44	11.68	NA	16.57	NA	9.81	NA	4.04
DNF	4.77	0.83	2.33	0.47	6.04	NA	6.32	NA	4.56	NA	1.58
SHR	14.01	1.73	4.68	0.42	11.33	NA	14.17	NA	18.48	NA	5.08
C3Gw	9.29	0.58	12.73	0.64	8.52	NA	9.96	NA	7.81	NA	1.90
C3Gk	5.84	0.36	12.31	0.62	8.16	NA	6.46	NA	10.38	NA	2.71
C3Gc	3.95	0.25	4.57	0.23	3.87	NA	5.08	NA	3.32	NA	0.68
C3C	7.66	0.71	16.34	0.82	10.90	NA	13.61	NA	7.86	NA	3.74
C4G	17.26	1.09	17.26	1.09	16.62	NA	17.26	NA	17.45	NA	0.32
C4C	3.23	0.30	3.23	0.30	3.47	NA	3.23	NA	3.16	NA	0.12

Note: the remote sensing-based products include MODIS v6.1 (Sulla-Menashe and Friedl, 2018), ESA CCI v2.0.8 (Harper et al., 2023), CGLS_LC100 v3 (Buchhorn et al., 2020), GLC_FCS30D (Zhang et al., 2024), and GLASS-GLC (Liu et al., 2020). Uncertainties (unc) for MODIS and ESA CCI PFT maps are derived from literature (see Methods). Uncertainties for other products are not available (NA).

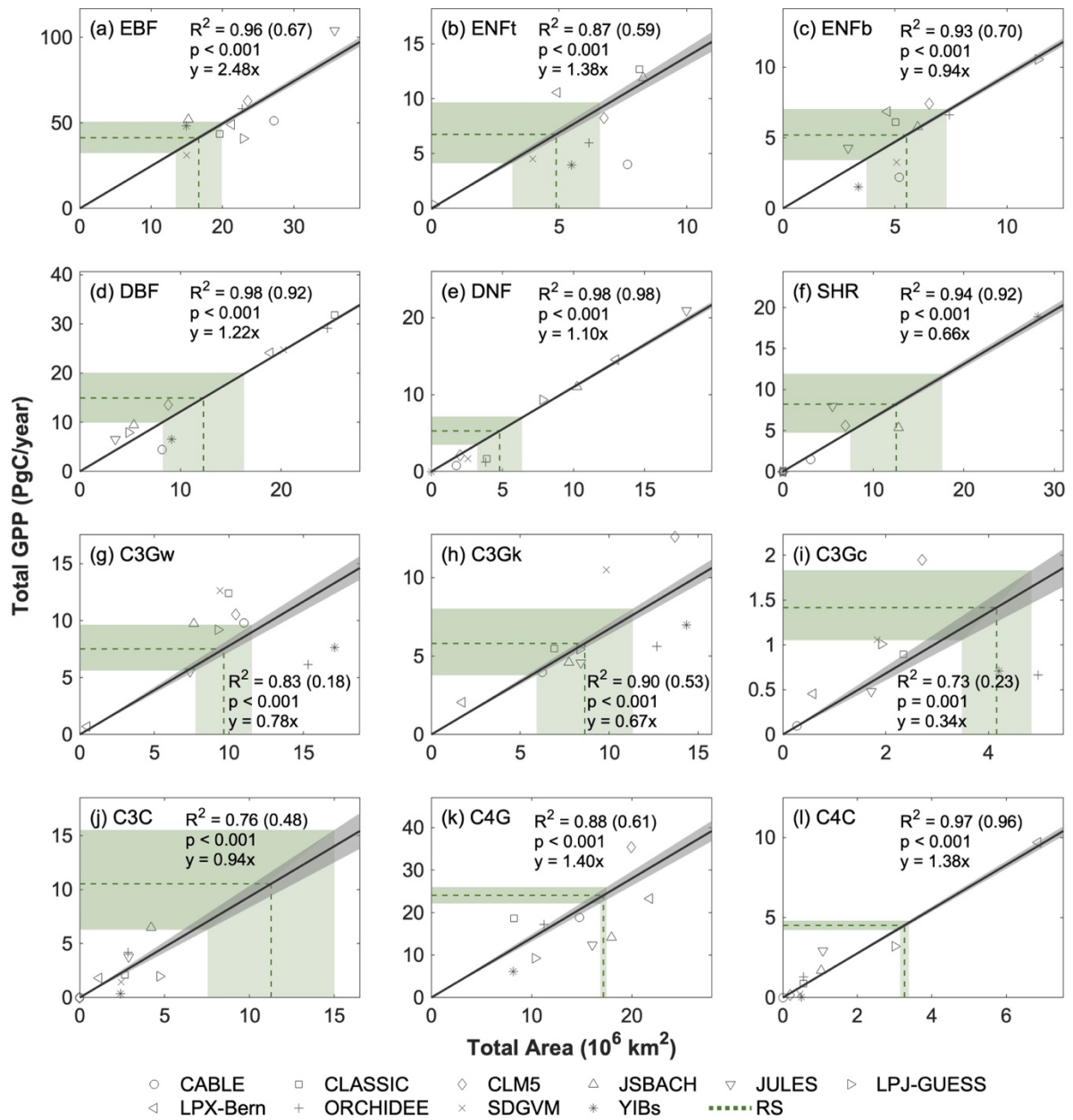


Figure 1. The relationships between total gross primary productivity (GPP) and total area for each PFT. The panels represent the relationships between GPP and area for (a) evergreen broadleaf forests (EBF), (b) temperate evergreen needleleaf forests (ENFt), (c) boreal evergreen needleleaf forests (ENFb), (d) deciduous broadleaf forests (DBF), (e) deciduous needleleaf forests (DNF), (f) shrubs (SHR), (g) warm C_3 grasses (C3Gw), (h) cool C_3 grasses (C3Gk), (i) cold C_3 grasses (C3Gc), (j) C_3 crops (C3C), (k) C_4 grasses (C4G), and (l) C_4 crops (C4C). The uncertainty of the relationship was quantified as the upper and lower boundary of the relationship through a bootstrapping method for the regression (see Methods). The regression lines were forced through (0,0) as in theory there would be no GPP if PFT area was zero. The strength of the relationships was assessed by the coefficient of determination (R^2). Meanwhile, we also provided R^2 for the regression lines without forcing through (0,0) in parentheses. The vertical green lines stand for the average PFT areas from five remote sensing-based (RS) PFT maps, including MODIS v6.1 (Sulla-Menashe and Friedl, 2018), ESA CCI v2.0.8 (Harper et al., 2023), CGLS_LC100 v3 (Buchhorn et al., 2020), GLC_FCS30D

(Zhang et al., 2024), and GLASS-GLC (Liu et al., 2020). The vertical shadings indicate the one standard deviation of the PFT area estimated by the five RS PFT maps. The horizontal green lines and shading intervals represent the constrained total GPP for each PFT using RS PFT areas.

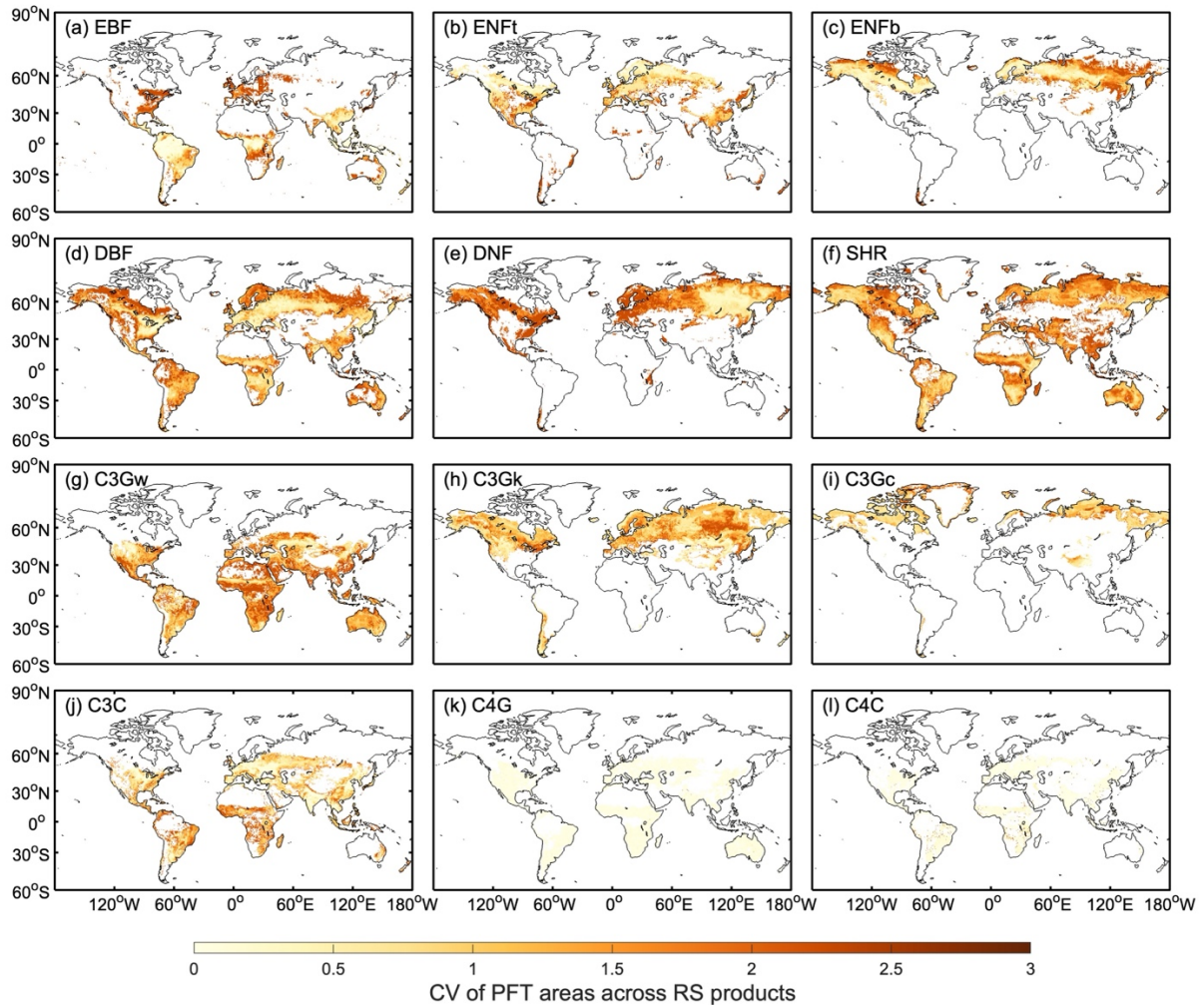


Figure 3. The difference in the distribution of PFTs from remote sensing-based (RS) products. The coefficient of variation (CV) for each PFT distribution across RS products is calculated as the ratio of the standard deviation to the mean PFT area, with higher CV values indicating greater uncertainties in PFT simulations. The abbreviations for PFTs and RS products correspond to those in Figure 1.

R1C3: Second, the authors may also clarify the potential limitation of emergent constraint in term of underlying mechanisms underlying mechanisms (i.e., Winkler et al., 2019) and the spatial/temporal scales mismatch between data and models (i.e., Winkler et al., 2019). Ideally or necessarily, the research efforts in reducing the uncertainty of land carbon sink should be aimed by both improving the model structure or processes (i.e., the emergently vegetation biome distribution) and the emergent constraint approach here. This also related to the constraint of ER in this study because the current model structure have very simple structure for modeling ER which are strongly coupled with GPP.

Response: We hope to clarify that in the last round of revision, we have stepped away from stating our method as emergent constraint – we understand that there is a debate within the community now about the definition of emergent constraint (please see our replies to reviewers in the last round), and we did not intend to engage more into that discussion. Therefore, we now refer to our method as a simple correlation-based approach. The relationship we identified, strongly demonstrate the key role of vegetation biogeography on land carbon cycle modelling, and motivate future carbon cycle studies to consider biogeography (and their associated structures and parameters, L286-290) as a main place for improvements.

R1C4: Third, the authors may also consider the potential effects of land usage impacts. Land use change likely have influenced the vegetation biome distribution more than climate change in term of spatial and temporal scales. Currently, it is not clear how the land usage change has been accounted for.

Response: Thank you. In this study, we focused on how inter-model difference in GPP modelling is influenced by inter-model difference in PFT distribution. We mainly used the S2 scenario estimates from TRENDY v9, which did not consider the effect of LUCC. Motivated by reviewer's comment, we used the S3 scenario simulations, which include LUCC impact, to redo the analysis. We found the strong relationships between total GPP and total area for each PFT persist in S3 scenario (Fig. S9), which include the impact of LUCC. The new analysis indicates that the strong PFT GPP and Area relationships exist at the global scale in all scenarios. We have included a statement in L371-374 to clarify that.

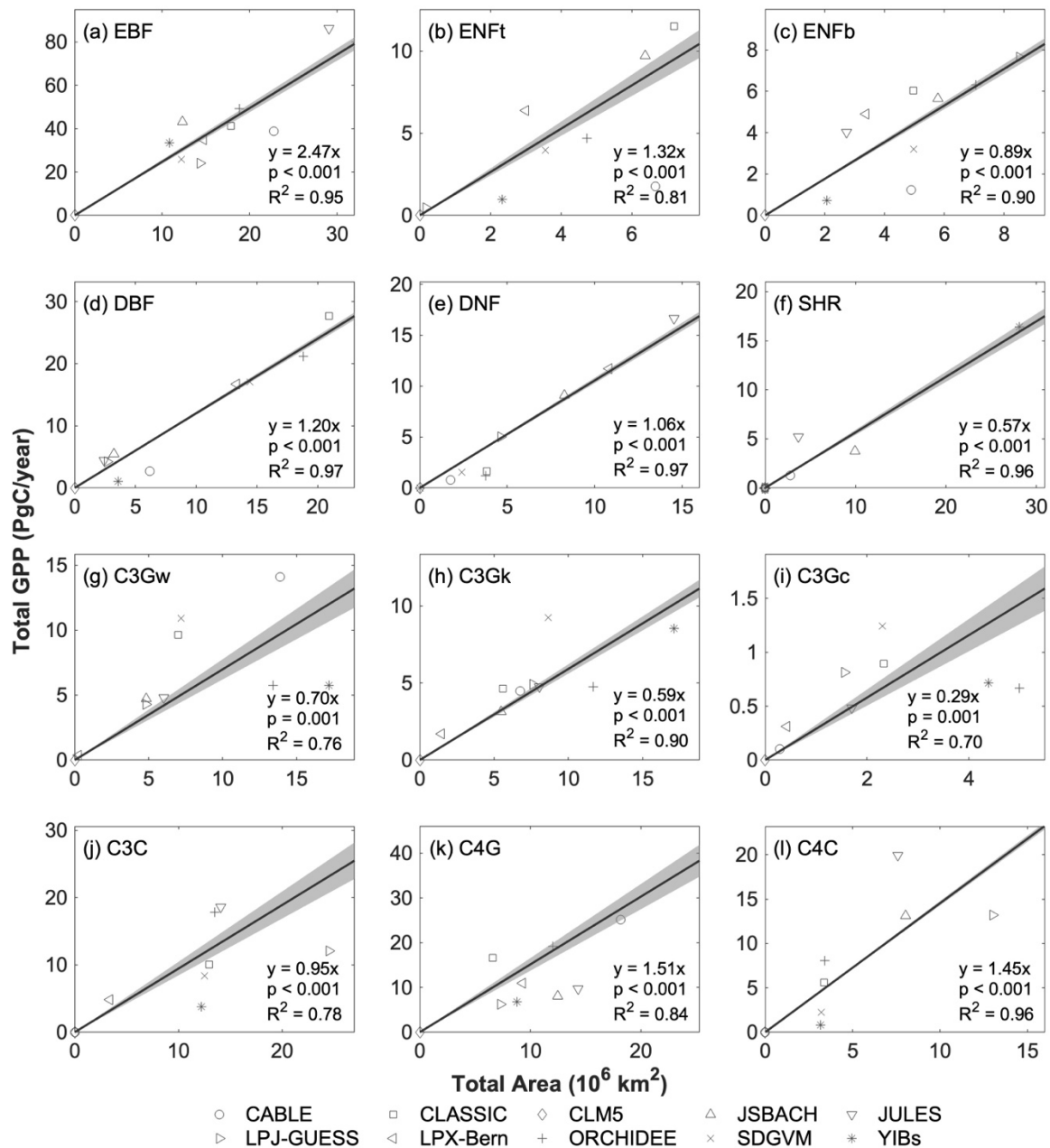


Figure S9. The relationships between total GPP and total area for each PFT. Here we presented the results using S3 scenario simulations from TRENDY v9.

R1C5: Since the paper is very well written with insightful discussion, I also have the following minor suggestion for the authors to consider

Abstract: may also consider reporting the results of ER

Response: Thank you for the positive comment on our discussion section. Following the suggestion, we have included the results about uncertainty reduced (i.e., approximately 75%) for ER using remote sensing-based PFT maps to further highlight the importance of biogeography for land carbon cycles (L25).

R1C6: Lines 44-45: it would be important to mention competition, i.e., by accounting for size or cohort based competition for resources (light, nutrients and water)

Response: Thanks. We have added this point in L48.

‘Considerable efforts have been made to identify the sources of uncertainty in models for land carbon cycle simulations, such as assessing climate forcing datasets and improving representations of ecological processes (e.g., thermal acclimation of photosynthesis, explicit representation of microbial processes, plant hydraulics, and competition for resources (Dantas De Paula et al., 2025)), and parameterizations (e.g., photosynthetic capacity and microbial carbon use efficiency) in process-based models.

R1C7: Introduction: may consider clarifying what the emergent constraint is and the limitation of emergent constraint.

Response: Please see our response to R1C3. Since the last version, we have stepped away from referring to our method as emergent constraint as the request of other reviewers, to avoid potential confusions.

R1C8: Figure 1: the color scheme would need to be improved. Also there are different definitions for vegetation biomes such as WWF vegetation biome. My knowledge is also that there is 1 km resolution of land cover maps <https://www.earthenv.org/landcover> I am not sure whether this is useful as an additional observational constraint.

Overall, the whole manuscript has focused on the constraint of GPP, but the results of ER was indeed also important for carbon sink. Thus, the authors may consider balancing this.

Response: Thank you for the suggestions. 1) We have updated Fig. 1 (now Fig. S2) by adopting a more bright colour scheme, following the suggestion from the reviewer.

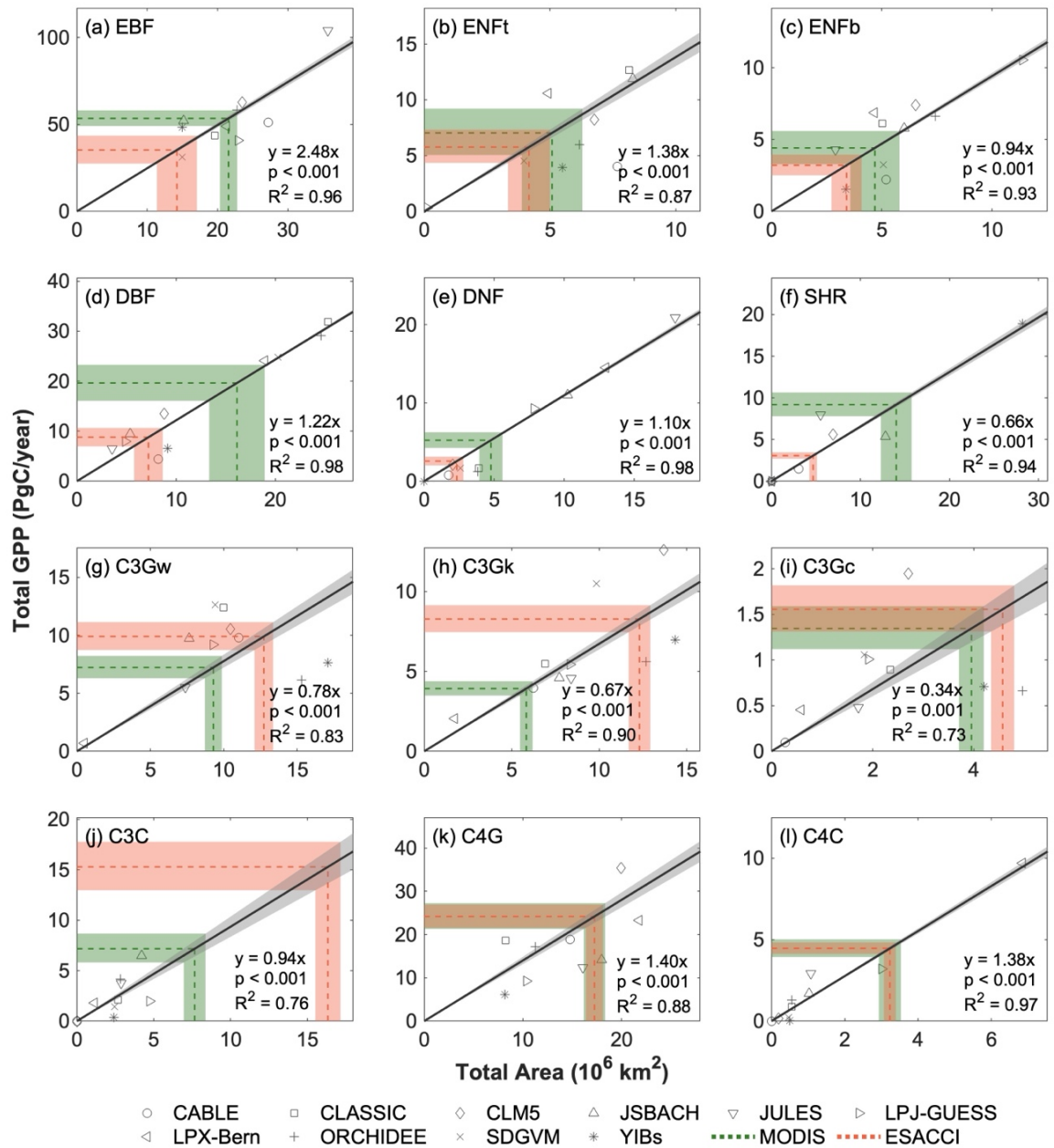


Figure S2. The relationships between total GPP and total area for each PFT. Here we presented the results similar to Figure. 1, but showing the PFT areas observed from MODIS and ESA CCI only. The vertical shadings indicate the one standard deviation of each PFT map derived from existing literature (see Methods). The horizontal green and red lines and shading intervals represent the constrained total GPP for each PFT using MODIS or ESA CCI PFT area, and the uncertainty in GPP propagated from the PFT area.

2) We agree with the reviewer on considering more RS PFT maps, such as the WWF vegetation biome classification and the land cover map that the reviewer mentioned. Please see our response to R1C2, in which we collected five widely-used global PFT maps, including the land cover that shared by the reviewer, to provide a more robust estimate of constrained GPP (Table S3 and Fig. 1).

Please kindly note that since WWF map does not adopt PFT classification and is not compatible with the PFT classifications in DGVMs, we ended up not using it. But we did include the one from the link that shared by the reviewer in our updated analysis.

3) As for ER constraint, we agree with the reviewer that ER is important as GPP in determining the carbon sink. However, the challenge is that only a limited number of DGVMs provide PFT-specific ER; thus, we do not have sufficient data to do the analysis on ER at the same depth as GPP. As an alternative, we have constrained global ER based on the strong relationship between global GPP and global ER. We highlighted that this aspect needs improvement in the DGVM outputs (L394-397).

R1C9: Figure 2: I am curious whether N deposition has been accounted for and how important it would be.

Response: That's an interesting question – broadly the reviewer is asking about the impact of N deposition on vegetation biogeography change. Unfortunately, the current TRENDY DGVMs do not design enough scenarios to allow assessment of its impact on PFT distribution. However, considering that N deposition primarily occurs at a regional scale – particularly in regions with strong GHG and aerosol emissions, such as China and India (Zhu et al., 2025), it is likely to affect PFTs at a regional level. There is some evidence suggesting N deposition benefits certain species than others (Wang et al., 2021), however, we do not see any observational evidence at the biome level. We would like to regard this as an open question for regional level PFT change, and can potentially examine it once more data is available.

R1C10: Discussion: as clarified, the authors may consider clarifying the limitation of emergent constraint and needs of improving model structure/processes for the mechanistically deriving/model the emergent properties of vegetation biome.

The section of discussion on structure (vegetation biome biogeographical distribution) vs biogeochemistry is interesting. It is also good to learn that the authors used emulator to disentangle their relative impacts. With this, the authors may consider elaborating a bit more for the implication of improving the model mechanisms in term of more research efforts on vegetation biome biogeographical distribution itself vs biogeochemistry.

Response: Please kindly refer to our response to R1C3 and R1C7, where we stated the reason for our withdrawal from using the term “emergent constraint” and clarified that we now refer to the relationship as a simple “GPP-area relationship”.

We appreciate the reviewer's suggestion on elaborating more for the implication of improving model in term of biogeography vs biogeochemistry. Based on our emulator analysis, we have expanded our discussion regarding future efforts on these two aspects (L286-290).

“Overall, the EM(DGVMs, MODIS) approach perhaps provides a more realistic and conservative estimate of biogeography's role in GPP modelling, and indicates that improvements in biogeography in GPP modelling should be accompanied by improvements in biogeochemistry.”

References

Winkler, A. J., Myneni, R. B. & Brovkin, V. Investigating the applicability of emergent constraints. *Earth Syst. Dyn.* 10, 501–523 (2019).

Reviewer #2 (Remarks to the Author):

Reviewer's comments

R2C1: This paper attempts to identify an observational constraint for dynamic global vegetation models (DGVMs) by establishing a linear regression relationship between plant function type (PFT) area and total gross primary productivity (GPP) for a given PFT. The authors use satellite-based observations to infer a plausible range of total global GPP per-PFT (GPP_{pft}). They correctly identified that land surface models with dynamic vegetation components are often deficient when it comes to simulating present-day biogeography, and that using these PFT observations to constrain total global GPP estimates can tighten the range of GPP uncertainty.

Response: We appreciate the comments from the reviewer and have endeavoured to improve our manuscript following their suggestions. Please kindly see the point-by-point response below, where the line numbers refer to those in the clean version of the revised manuscript.

R2C2: I have serious concerns, however that the linear regression on which this study is based is fundamentally flawed because both variables used in the regression have a strong dependence on PFT area. While I broadly agree with the conclusion that observed global area of a PFT can be used to refine global GPP estimates derived from DGVMs, I have serious reservations about the approach taken here. My criticism boils down to the fact that in this regression relationship, the independent variable (x-axis; PFT area) is also a central component of the dependent variable (y-axis; total global GPP_{pft}), and therefore the resulting strong relationships shown for each PFT are not valid because it is essentially using PFT area to predict a variable that is strongly dependent on PFT area. In fact, the only reason why there is variation in the relationship (and not a perfect 1:1 line) is because there is variation in the per-model and per-grid cell GPP_{pft} rate per unit area. In order to derive total global GPP_{pft}, I assume the authors would have had to multiply the raw land surface model output GPP_{pft} rate per unit area (likely in units similar to gC m⁻² yr⁻¹) by the grid cell PFT area, and then sum all grid cells (possibly also making a multi-year mean of the timeseries). Similarly, converting model output of PFT fraction to a total land area for a given PFT (Area_{pft}) would involve multiplying the per grid cell fraction by each grid cell's area (giving the exact value used in the GPP calculation in the previous sentence), and then summing all grid cells. If you take out the area dependence from the simple GPP calculation, you are left with the GPP_{pft} rate again per unit area. Therefore, I think the linear regression on which the emergent constraint is based is flawed.

Response: We agree with the reviewer about the description of the calculation of PFT total area and total GPP, and we have used a similar approach in our analysis. We sensed, however, that there might be a misunderstanding regarding the interpretation of our results, which we clarify below.

For a given PFT in a given DGVM, the total GPP of the PFT can be simplified as:

$$\text{GPP}_{\text{pft}} = \text{A}_{\text{leaf}} * \text{LAI} * \text{Area} \quad (1)$$

where Aleaf is the average photosynthesis rate per leaf area ($\text{g/m}^2/\text{yr}$), LAI is the leaf area index and Area is the total ground area for that PFT. Aleaf*LAI is the same as GPP rate that the reviewer mentioned.

Based on equation (1), Area is certainly an important variable for GPP estimation (as noted by the reviewer). However, it does not mean a strong relationship between GPP and Area is a given. In fact, Aleaf and LAI variation between models are more often regarded as critical to GPP modelling in most DGVM uncertainty analysis, while the variation in Area is less noticed and considered to be not a main reason for inter-model variation in GPP. Our result directly presented that Area is a first-order factor controlling the GPPpft, and the impact of Area is stronger than Aleaf and LAI. We have provided the analysis in the last round of review to support the point (now Fig. S8a).

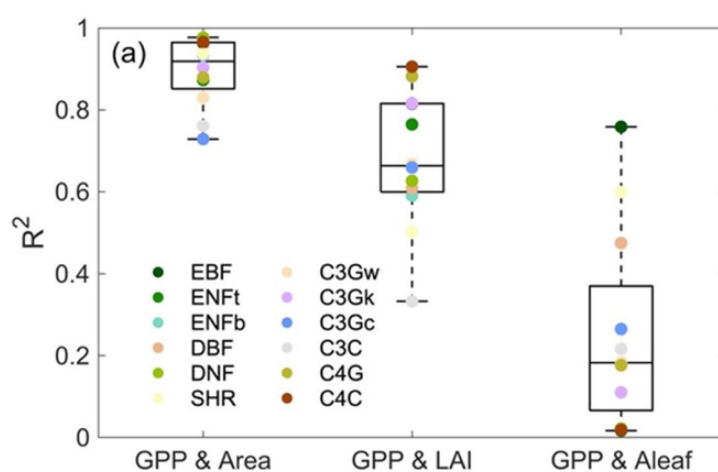


Figure S8a. Comparison of the relationships between carbon cycle components and PFT characteristics. (a) The coefficient of determination (R^2) for the relationships between PFT total GPP and PFT area (GPP & Area), or leaf area index (GPP & LAI) or leaf photosynthesis rate (GPP & Aleaf). The R^2 is the coefficient of determination for the relationships, with higher values indicate a stronger correlation. The abbreviations for PFTs correspond to those in Figure 1.

By presenting the relationship between GPPpft and Area across models, our study offers a way to assess the importance of PFT area to the uncertainty (i.e., inter-model difference) in GPP modelling. We understand that the reviewer had another question on the variation in GPP rate within PFT which is more elaborated in R2C15. Please refer to our reply to R2C15 below.

R2C3: Having said that, I do think there is value in highlighting that observations of PFT area can be used to constrain the simulated GPP of models in order to gain a more accurate measure of present-day GPP and consequently atmosphere-land carbon fluxes. I just don't think that it can be done in the way that the authors have proposed. I believe that the authors would learn more about the models and the plausible constraints on GPP by looking in detail at the locations where PFT fractions match observations, and those where PFT fractions do

not. I would also encourage the authors to examine other carbon-cycle emergent constraint studies and consider the definition presented in the Cox et al. (2019) paper:

“The emergent constraint technique involves using the full ensemble of models to find an across-ensemble relationship between an observable feature of the Earth System (such as a trend, interannual variation or change in seasonality) and an uncertain aspect of the future.”

I would argue that while PFT fraction is an observable feature of the Earth System (and GPP the uncertain aspect of the future), I don't think that it provides a trend, interannual variation or change in seasonality, in the same way that other emergent constraint studies have. See for example, Cox et al. (2013) who found a strong linear relationship between the sensitivity of tropical land carbon storage to warming and the sensitivity of the annual growth rate of atmospheric CO₂ to tropical temperature anomalies.

Response: Thank you for acknowledging the value in our study. We fully agree with the reviewer that there is a need to assess where the model PFT distribution is simulated differently. We have provided an analysis (updated Fig. S4, S5, S6) and a full discussion on that issue (L196-208). Also, we compared five widely-used PFT maps from remote sensing (RS), and quantified their disagreement over the world (updated Fig. 3). We have identified the key regions and key PFTs that needs improvement for mapping (L212-220).

We also hope to clarify that we have stepped away from referring to our method as “emergent constraint” in the last round, following the suggestion from the reviewer. We have now referred to our results as a “strong relationship between PFT GPP and PFT Area”.

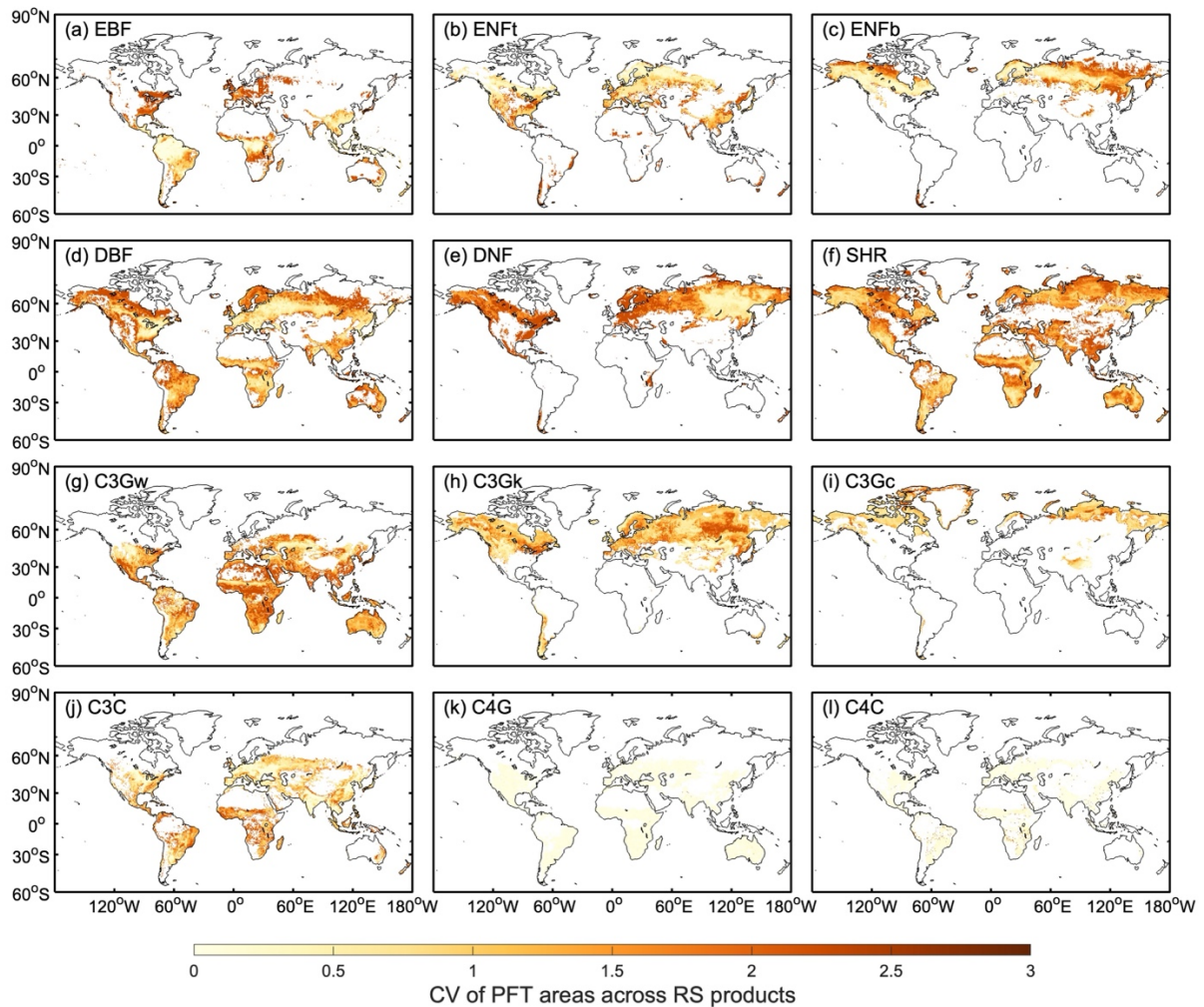


Figure 3. The difference in the distribution of PFTs from remote sensing-based (RS) products. The coefficient of variation (CV) for each PFT distribution across RS products is calculated as the ratio of the standard deviation to the mean PFT area, with higher CV values indicating greater uncertainties in PFT simulations. The abbreviations for PFTs and RS products correspond to those in Figure 1.

Other Comments

R2C4: Line 1: Title. I'd suggest reviewing the title, as I think "main source of uncertainty" implies that you've compared biogeography to other sources of uncertainty in the land carbon cycle.

Response: Thanks for your suggestion. Our full title conveys that "biogeography is a main source of uncertainty". We understand biogeography is not "THE" main source of uncertainty since other biogeochemical factors are also important, but we believe our results are sufficient to indicate that the inter-model difference in PFT distribution substantially contributed to the inter-model difference in land carbon cycle modelling, and should be regarded as "a" main source of uncertainty.

R2C5: Line 25: Worth making clear that you've reduced the spread of land carbon estimates for the present day, or does it also cover the future?

Response: Our analysis is based on the estimates of DGVMs in the past twenty years. To clarify, we have updated this sentence to ‘The dependence allowed us to reduce the spread of land carbon cycle estimates by approximately 75% using remote sensing-based PFT maps.’ We have stepped away from stating our approach as emergent constraint, given the term perhaps implies a future prediction.

R2C6: Line 26: Again, emphasise that you’re talking about historical changes

Response: We have updated our statement to ‘We further found that $56 \pm 21\%$ of the climate-driven changes in global photosynthesis modelled by DGVMs were caused by changes in PFT distribution in the last two decades’, following the suggestion from the reviewer.

R2C7: Line 33: the term “offsetting” tends to be used in carbon trading. I’d suggest being more quantitative here – perhaps cite the fraction of anthropogenic CO₂ that the land absorbs?

Response: We have replaced ‘offsetting’ with ‘absorbing’ as suggested, and revised the sentence to ‘by absorbing approximately 25% of anthropogenic CO₂ emissions (Friedlingstein et al., 2023)’.

R2C8: Line 54: Limitations could also include consistency of satellite timeseries, particularly as we’re interested in the response of the C cycle to increasing CO₂ emissions

Response: We have added “lack of or inconsistent consideration of CO₂ fertilization effect (De Kauwe et al., 2016; Keenan et al., 2023)”, as suggested by the reviewer.

R2C9: Line 58-59: See line 1 comment.

Response: Please kindly refer to our response to R2C2, where we clarified the key reasoning in our research.

R2C10: Line 76: Maybe clearer to say “global” GPP here

Response: We have updated our statement to “We aim to quantify the dependence of PFT total GPP on PFT distribution and quantify the contribution of vegetation biogeography to the uncertainties in global GPP and ER estimation (see Methods)” for clarification.

R2C11: Line 82: Would perhaps be better to assess how improved estimates of GPP affect the uncertainty in land carbon fluxes, and improve our ability to close the C budget over this period

Response: This is a really good point, and we did intend to conduct our analysis for the full C budget. However, a key limitation is that in TRENDY project output, many DGVMs only provide PFT-specific GPP but not ER or NEE, making it challenging to apply our approach to constrain ER or NEE for each PFT.

Alternatively, we used constrained estimate of global GPP to narrow down global ER, based on the relationship between their global values across models (Fig. 2c). We ended up with a constrained global GPP of 135.66 ± 6.54 PgC/year and a constrained ER of 133.01 ± 6.53 PgC/year. We were unable to obtain a constrained NEE, as inter-model NEE is not correlated with inter-model GPP or ER. We have expanded our discussion (L394-397) to clarify this issue.

R2C12: Line 87: Is the PFT total area also in km² / year?

Response: The PFT total area reported here represents the average area over the period of 2001–2019, rather than an annual value or rate of change. Therefore, its unit is km², not km²/year. We hope to clarify that we are mostly interested in inter-model difference in this study. We have clarified this in L90.

R2C13: Line 90-92: The premise of the whole paper hinges on this statement – see main reviewer comments.

Response: Please kindly refer to our response to R2C2 and R2C15.

R2C14: Line 93: Why didn't the authors split EBF into tropical and temperate in a similar way to grasses? There would likely be a huge range in the GPP rate within the EBF PFT.

Response: We did not split EBF into tropical and temperate EBF as most DGVMs regard EBF as one PFT and do not separate it into tropical and temperate, and most EBF is actually distributed in the tropics. We understand there would be a strong variation in GPP rate (g/m²/year) within one PFT, including EBF, however, analyzing such within-model variation is not the objective of our study. Our study focused on the *inter-model* discrepancy in GPP estimates, and therefore the *inter-model* difference in GPP rate and Area, not the spatial or temporal variations in GPP in *one model*.

R2C15: Line 93-101: How do these values relate to the annual rate of GPP that would have come directly out of the models? Within a given PFT distribution, there would be a range of GPP rates (measured in similar units such as gC/year/sqm), so how are these rates different from just using the observed PFT distribution maps and taking a spatial mean and standard deviation of the modelled GPP rate?

Response: We apologize for the confusion – we noticed that a key point of misunderstanding is that we are discussing *inter-model* difference in GPP while the reviewer seemed to have concerns on the GPP rate variations within one PFT.

Unfortunately, we cannot do “how are these rates different from just using the observed PFT distribution maps and taking a spatial mean and standard deviation of the modelled GPP rate” because DGVMs might not have simulations for a given PFT at places where remote sensing (RS) observations report there should exist that PFT. Even when there is overlap between DGVM PFT and RS PFT distributions, DGVM might simulate a very biased vegetation status (such as LAI) for a given PFT in that region, and consequently result in a biased GPP rate. Therefore, it would be very challenging to conduct the analysis from the pixel level (pixel level GPP rate for each PFT). Instead, our analysis is conducted at the biome scale (total GPP and total area for each PFT).

To avoid misunderstanding of the results, we have rephrased the statement to ‘For each of the twelve PFT types on the land surface, we observed a strong and significant relationship ($p < 0.001$) between GPP and area across models (Fig. 1). The strong relationships indicate that the inter-DGVMs discrepancy of total GPP of each PFT is largely caused by the difference

in the prescribed or simulated PFT area of DGVMs, and that DGVMs simulate a similar per-area GPP on average for each PFT (i.e., the slopes of the GPP-area relationships in Fig. 1).'

R2C16: Line 101-103: This statement is probably for the discussion, and needs referencing / supporting with evidence from study

Response: We have now moved the statement to discussion (L307-309) and added references (Hu et al., 2018; Rogers et al., 2017), following the suggestion from the reviewer.

R2C17: Line 177: This relationship is bound to be strong if both variables co-vary strongly with area

Response: We appreciate your comment; however, we believe that the strong correlation between global GPP and global ER is not merely driven by their covariance with area. As we clarified in our response to R2C2, the strong relationship between PFT total GPP and total area across DGVMs is not a given, nor is the relationship between ER and area. The observed strong relationship between global GPP and global ER, indicates a strong ecological linkage between carbon uptake (GPP) and carbon release (ER), though in reality the relationship might not be that strong, considering the different factors controlling GPP and ER. The strong relationship between global GPP and global ER across models is likely due to a carbon sink centric view adopted by DGVMs, where the input of the carbon through photosynthesis largely determined the magnitude of ER (Fatichi et al., 2019), particularly over the long term.

R2C18: Line 229: Emphasise that it is PFT-specific total GPP

Response: We have updated the statement to "PFT-specific total GPP" in L232.

R2C19: Line 280-281: I don't think you can conclude this from your results. I would argue that the higher uncertainty reduction in the approach presented in figs 1 & 2 is due to the over-dependence on area in the total GPP vs PFT area relationship

Response: Please kindly refer our response to R2C2 regarding the dependence of PFT total GPP on PFT total area.

R2C20: Line 282-283: Again, I don't think you can conclude this from your results. There is a lot of literature that shows parameter tuning from detailed site-specific observations - see for example Harper et al who used plant trait information to tune JULES, in places where the PFT fractions were very well documented. <https://gmd.copernicus.org/articles/9/2415/2016/gmd-9-2415-2016.pdf>

Response: Thank you for sharing the study with us. We have removed the statement as we found it challenging to make a generalised statement on the strategies of tuning parameters for different DGVMs.

R2C21: Line 331: I think it's important to be specific here - your study attempts to understand and reduce uncertainty in the *historical* to present-day carbon cycle. It highlights that DGVMs have work to do to improve the modelled spatial distribution of the their PFTs.

Response: Thank you for your suggestion. We have stated that we found the *inter-model* difference in land carbon cycle simulation is related to inter-model difference in PFT

distribution, and that issue should persist regardless of whether we use DGVM simulations for past, present or future, though our analysis used DGVM simulations for the past twenty years. We have stated in our method we used TRENDY v9 simulations from 2000 to 2019 to be compatible with remote sensing-based PFT maps.

R2C22: Line 350-351: Worth acknowledging (apologies if I've missed it) that in using the S2 scenario, we would expect DGVMs to over-estimate the fraction of tropical broadleaf evergreen tree (which is the PFT class that normally has the highest contribution to global GPP)

Response: Thanks for pointing this out. We have revised this sentence to acknowledge we used S2 scenario in our main analysis. Moreover, by doing additional analysis using S3 scenario, which further includes the impact of land use and land cover change, we have confirmed that our main findings remain robust (Fig. S9). We clarified this point in L370-374,

"We mainly used the simulations from the S2 scenario to examine the relationship between PFT-specific total GPP and area. Changing the simulations to other scenarios, such as S3, would not affect the relationship between PFT total GPP and PFT area across models (Fig. S9), however we would like to exclude the consideration of land use and land cover change in this analysis."

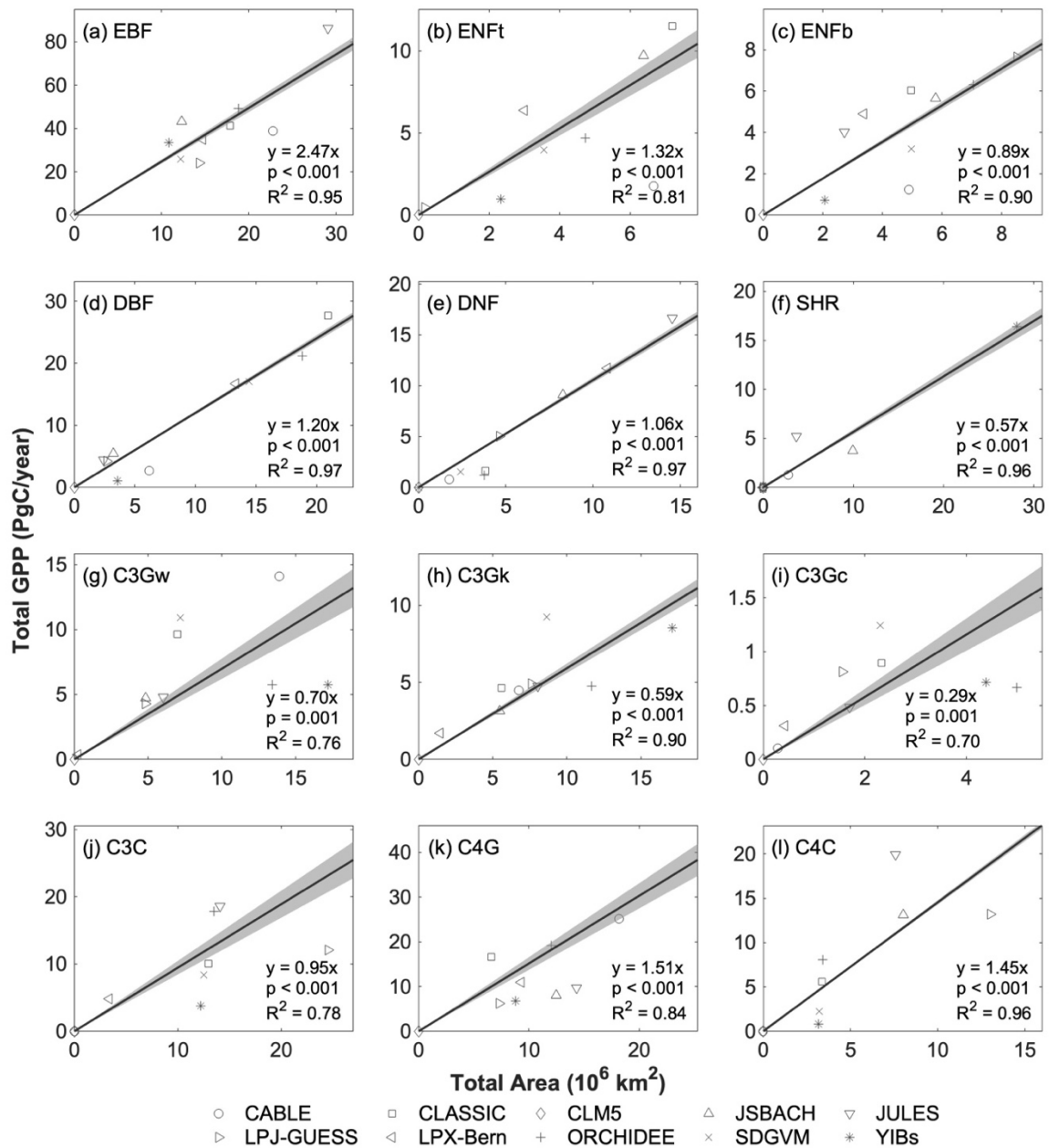


Figure S9. The relationships between total GPP and total area for each PFT. Here we presented the results using S3 scenario simulations from TRENDY v9.

R2C23: Line 359: See earlier comment on Tropical vs Temperate EBF

Response: Please kindly refer to our response to R2C14.

R2C24: Line 378-379: Maybe a better description of the metric is need here, given the $>1200\text{C}$ criteria, I would assume it is rather something like accumulated temperature during the growing season?

Response: We apologize for the unclear description here. Based on Prentice et al (1992), we have revised the statement to 'an accumulated growing degree-days above 5°C (GDD_5) exceeding 1200°C '.

R2C25: Line 417-418: Is this relationship on an annual timestep?

Response: We examined this relationship based on multi-year (2001-2019) average GPP_{pft} and $Area_{pft}$. We have clarified this by revising this sentence to 'We obtained the multi-year average total GPP and total area for each PFT over the period 2001-2019, and examined the relationships between them'.

R2C26: Line 419: I think I'm right in saying that β_i is the GPP rate per unit area here. In that case you would also need a sigma to sum across grid cells.

Response: As clarified in our response to R2C2, β_i represents the dependence of total GPP on total area for i^{th} PFT ($i = 1, 2, 3, \dots, 12$, representing EBF, ENFt, ENFb, ..., C3C) across models (i.e., the slopes of the total GPP – total area relationships for each PFT in Fig. 1). We already summed pixel level values to the biome scale when calculating PFT total GPP (unit: PgC/year) and PFT total area (km^2), separately. Therefore, an additional summation (sigma) across grid cells is not required. Please also kindly refer to our answers to R2C2 and R2C15, where we hope to clarify this point of confusion in the review process.

Reviewer #3 (Remarks to the Author):

R3C1: Zhao et al present a method for constraining total global GPP and also to explore relationships between change in GPP over time with changes in PFT area. To constrain global GPP the authors define linear relationships across models between total global PFT specific GPP and PFT area and then constrain the range in global GPP values based on independent (though see caveat on that below) datasets of PFT area. In so doing, the authors highlight the importance of PFT area in modeled global GPP uncertainties. They also (for one independent dataset) constrain the mean global GPP to a higher value than is often reported, but which is more in line with other independent total GPP estimates.

I served as a reviewer of the original version of this manuscript. I appreciate the extensive additional work the authors have done to revise the manuscript and to add extra, very involved analyses to address some of the concerns.

In my original review I emphasized that the goal of this study, to highlight the importance of differences in PFT area on uncertainties modeled GPP estimates, was much needed. I still think that, and furthermore I think the changes the authors have made to the paper have enhanced that message. I am very supportive of seeing this paper published.

In the original review I had concerns about the approach that have mostly been addressed. The difference in framing alone helps me to better digest and have greater confidence in the results and conclusions of this study. However, I still see some caveats that affect the strength of the conclusions drawn. I highlight these below, followed by key suggestions to address what I see as the remaining caveats. Following this general comments section I provide additional suggestions for several of the methods and results section before finishing with some minor comments.

Please forgive any repetition in the sections below. I have reviewed this paper over the course of the week because like this paper and I think it's important and therefore I wanted to provide a thorough review. This required me thinking through some parts of the new analyses and responses to reviewers and took longer than I expected. I have tried to tidy up my comments but in a rush to get this review to you I may have missed some repetition. Apologies for that. I would be happy to discuss any of these comments further, either via the journal editors or directly via email.

-- Natasha MacBean (nmacbean@uwo.ca)

Response: We thank the reviewers 3 and 4 very much for their time and effort in reviewing our manuscript and for their thoughtful comments. We are glad to learn that we have resolved most of their concerns and they are supportive of seeing this paper published. The reviewers provide some very in-depth thoughts, and we would be happy to clarify, and in some cases, explore with them below.

R3C2: Remaining caveats with the approach used and therefore strength of the conclusions drawn:

- **R3C2.1:** I think the understanding gained from the linear relationships between total GPP and PFT area are more nuanced than the authors lay out in their paper (more details below).
- **R3C2.2:** I think the uncertainties on the PFT maps are too narrow and therefore the large reduction in total GPP spread obtained by using the PFT area from these two maps is not realistic (more details in methods and results below).
- **R3C2.3:** The fact that the two different LC maps result in widely different mean total GPP estimates, even with (what I think are) too narrow uncertainty bounds, highlights the issue that the reduction in spread of total GPP values from constraining with satellite derived PFT area (as shown in Figure 2) is unrealistically high. Therefore the strength put on this statement – the reduction of 86% – just does not seem to be reliable.
- **R3C2.4:** There is a circularity in using the ESA CCI map at least, because some models are using that map to prescribe PFTs.

Suggestions to address these caveats:

- I will let the authors digest what I have said about the R2 values (especially those not constrained to go through zero), the strength of the relationship between total GPP and PFT area, and what (I think) that says about our understanding of differences in PFT representation across models. But I think that there needs to be a more nuanced description and discussion of those results (and this section is currently very short). I would not place as much strength in the statement in lines 90-92 that “The strong relationships indicate that the total GPP of each PFT across DGVMs is largely dependent on its prescribed or simulated area, and that DGVMs simulate a similar per-area GPP for each PFT”.
- I would like to see more realistic, wider uncertainties for each of the PFT products that included LC mapping uncertainties propagated through cross-walking uncertainties to the PFT maps (unless I’m missing this here – see more detailed comments in methods below), and therefore wider ranges of global total GPP (and a lower reduction on the total GPP spread) as a result. However, I would not wish that task of propagating cross-walking uncertainties to PFT maps on anyone! It was difficult enough in the Hartley et al. (2017) paper and we only explored some of the possible uncertainty “pathways” to demonstrate the issue – not the full range of uncertainties (which again leads me to think cross-walking uncertainties has not been accounted for here). So at the very least, I would like to see a clear statement and discussion that these error bars in Fig. 1 do not incorporate all sources of uncertainty in the final PFT maps, especially not in relation to the highly subjective cross-walking uncertainties, bioclimatic limits and somewhat arbitrary phenology thresholds used to link between LC class and PFT and to differentiate PFTs of different biomes, and therefore the real range could be much higher. Still, and this is a key point, even acknowledging wider uncertainties on the PFT maps, this could still constrain GPP to a higher value than is currently assumed (120-130PgC) – for MODIS at least.
- Further to the previous point, I would like to see more discussion of all the sources of uncertainties in satellite derived PFT maps and what we as a community could do about them.

This is mentioned in the discussion, but if there is space (I guess not) I would like to see more. Simply because the authors clearly do show that PFT area is an important source of uncertainty – that much is without a doubt for me. So which product is the best to use to prescribe PFTs in models, or to evaluate dynamic vegetation models (or to use as a constraint in the approach used in this study)? Do we just need an extensive worldwide campaign of validation of PFT maps – perhaps using citizen science? What else can be done? Examples of using ancillary datasets of tree cover fraction, as has been done with ESA CCI, sound good to me, but I am sure there could be other approaches. Perhaps we could explore the idea of optical plant functional types (but see minor comment below in response to R1C3). Also we could use the approach you have taken with emulators to fully explore more extensive PFT map uncertainties.

Response: We appreciate the comments from the reviewer and their suggestions on how to address the comments. Following their suggestions, we have done the following analyses.

R3C2.1 – We agree that the relationship of total PFT area and total GPP is not always strong. We therefore expanded our discussion in L303-307 to clarify that, and highlight some PFTs that have relatively weak correlations (such as C_3 grasslands). For those PFTs without a strong relationship, we expect their GPP uncertainty is more contributed by the biogeochemical process in DGVMs.

R3C2.2 – We fully agree with the reviewer regarding the uncertainty in remote sensing-based (RS) PFT maps, it is somewhat disheartening to see that after a few decades of development, we still cannot get a reliable/consistent RS PFT map. The uncertainty in RS PFT maps comes either from their own production process (e.g., cross-walking) or from the difference across products in their classification criteria. To fully assess the uncertainty, we now collected five widely-used RS PFT maps and listed their own and inter-product uncertainty. We are now using inter-product PFT uncertainty in the main analysis since many products actually do not have “own production uncertainty” reported (Table S3). We added a full discussion on the issue (L334-344).

After some consideration, we have decided to update our main results by using the mean of five RS PFT maps and the inter-product uncertainty in Figure 1. Using the ensemble of RS PFT maps, we still find a 75% decrease in GPP uncertainty. Though this reduction is less than when only using the MODIS PFT and ESA CCI maps (86%), but is still very substantial and supports our main conclusion that vegetation biogeography is a main source of land carbon cycle modelling.

We appreciate the reviewer's understanding that it is beyond our effort to discuss RS PFTs mapping in this study, as our key is to demonstrate the strong role of PFT area in influencing carbon cycle modelling, which is largely neglected by previous studies. We do intend to further explore PFT mapping in future and would like to discuss with the reviewers through other channels.

Table S3. The PFT area estimates across five remote sensing-based products (unit: 10^6 km^2). Only MODIS and ESA CCI PFT maps have reported part of their own uncertainty. σ refers to inter-product uncertainty (i.e., one standard deviation of the area from five products).

PFTs	MODIS	unc	ESA CCI	unc	CGLS_ LC100	unc	GLC_ FCS30D	unc	GLASS- GLC	unc	σ
EBF	21.56	1.23	14.23	2.85	18.30	NA	15.30	NA	13.97	NA	3.23
ENFt	5.08	1.21	4.17	0.83	7.32	NA	5.24	NA	2.61	NA	1.72
ENFb	4.69	1.11	3.41	0.68	6.20	NA	5.13	NA	8.17	NA	1.79
DBF	16.13	2.79	7.21	1.44	11.68	NA	16.57	NA	9.81	NA	4.04
DNF	4.77	0.83	2.33	0.47	6.04	NA	6.32	NA	4.56	NA	1.58
SHR	14.01	1.73	4.68	0.42	11.33	NA	14.17	NA	18.48	NA	5.08
C3Gw	9.29	0.58	12.73	0.64	8.52	NA	9.96	NA	7.81	NA	1.90
C3Gk	5.84	0.36	12.31	0.62	8.16	NA	6.46	NA	10.38	NA	2.71
C3Gc	3.95	0.25	4.57	0.23	3.87	NA	5.08	NA	3.32	NA	0.68
C3C	7.66	0.71	16.34	0.82	10.90	NA	13.61	NA	7.86	NA	3.74
C4G	17.26	1.09	17.26	1.09	16.62	NA	17.26	NA	17.45	NA	0.32
C4C	3.23	0.30	3.23	0.30	3.47	NA	3.23	NA	3.16	NA	0.12

Note: the remote sensing-based products include MODIS v6.1 (Sulla-Menashe and Friedl, 2018), ESA CCI v2.0.8 (Harper et al., 2023), CGLS_LC100 v3 (Buchhorn et al., 2020), GLC_FCS30D (Zhang et al., 2024), and GLASS-GLC (Liu et al., 2020). Uncertainties (unc) for MODIS and ESA CCI PFT maps are derived from literature (see Methods). Uncertainties for other products are not available (NA).

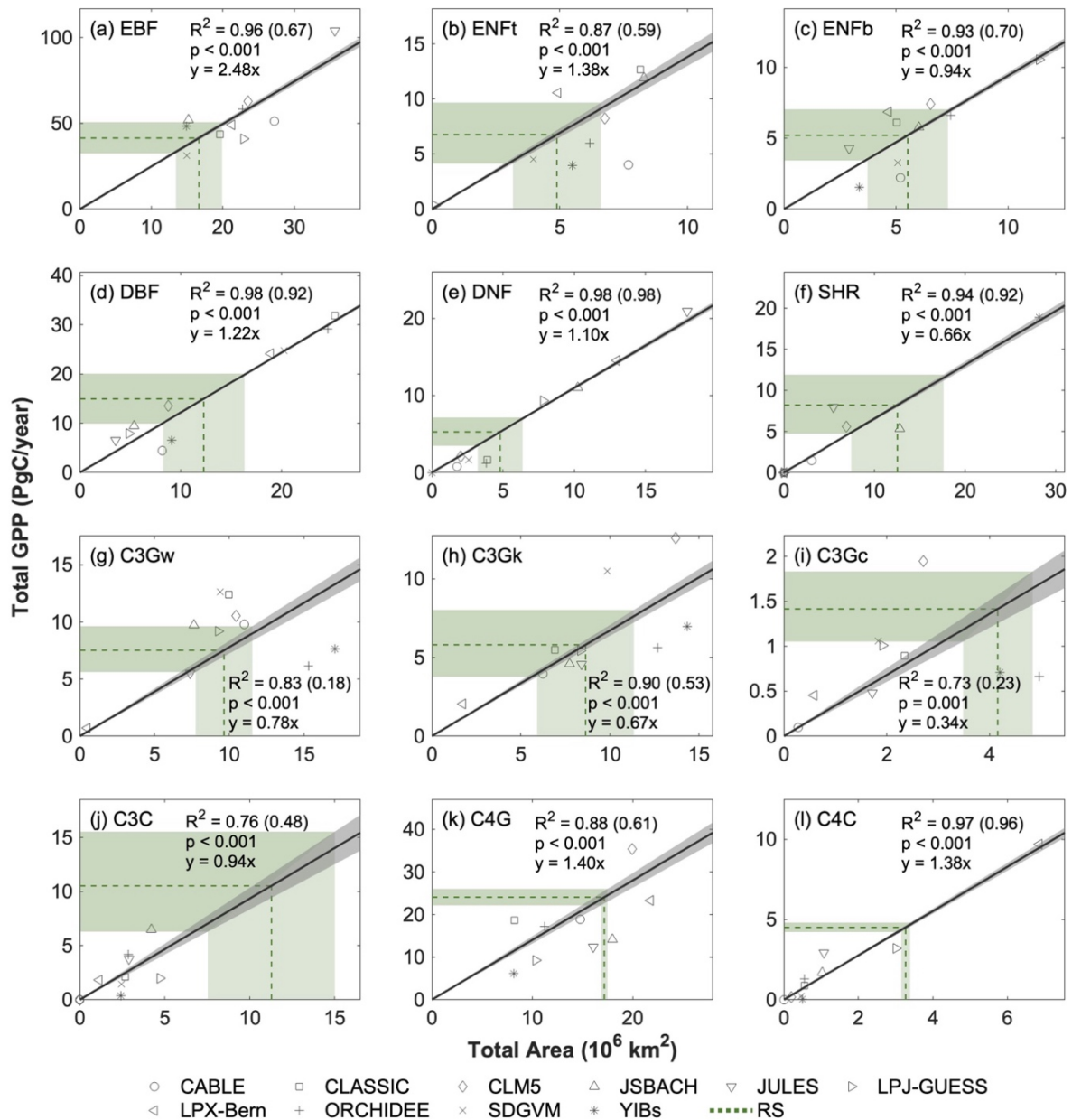


Figure 1. The relationships between total gross primary productivity (GPP) and total area for each PFT. The panels represent the relationships between GPP and area for (a) evergreen broadleaf forests (EBF), (b) temperate evergreen needleleaf forests (ENFt), (c) boreal evergreen needleleaf forests (ENFb), (d) deciduous broadleaf forests (DBF), (e) deciduous needleleaf forests (DNF), (f) shrubs (SHR), (g) warm C_3 grasses (C3Gw), (h) cool C_3 grasses (C3Gk), (i) cold C_3 grasses (C3Gc), (j) C_3 crops (C3C), (k) C_4 grasses (C4G), and (l) C_4 crops (C4C). The uncertainty of the relationship was quantified as the upper and lower boundary of the relationship through a bootstrapping method for the regression (see Methods). The regression lines were forced through (0,0) as in theory there would be no GPP if PFT area was zero. The strength of the relationships was assessed by the coefficient of determination (R^2). Meanwhile, we also provided R^2 for the regression lines without forcing through (0,0) in parentheses. The vertical green lines stand for the average PFT areas from five remote sensing-based (RS) PFT maps, including MODIS v6.1 (Sulla-Menashe and Friedl, 2018), ESA CCI v2.0.8 (Harper et al., 2023), CGLS_LC100 v3 (Buchhorn et al., 2020), GLC_FCS30D

indicating greater uncertainties in PFT simulations. The abbreviations for PFTs and RS products correspond to those in Figure 1.

R3C3: Methods – DGVMs

It's not clear from the text, unless I missed it, that some of these models have prescribed, static PFT fractions and some have dynamic vegetation. It reads in the methods text as if all the models have dynamic vegetation, but that's not the case is it? In fact, from reading the first section of the results it becomes clear that some models have prescribed and some simulate their dynamic vegetation. In this format of paper people may read the methods after the results, but for clarity I would add the information about whether the models have prescribed or simulated vegetation fractions to the DGVM methods section (and/or maybe Table S1)?

Following on from this point – if most models have prescribed vegetation then the statement that parameters feedback on PFT distributions is not correct for those models that do not have dynamic vegetation enabled, and this confused the additional results that are presented in Figure S5. More on this below.

Response: We have clarified in Method section and updated Table S1 to indicate whether DGVMs enabled dynamic biogeography modelling or not. In short, most DGVMs prescribed static PFT fractions, except for JULES, LPJ-GUESS, LPX-Bern, and SDGVM.

Table S1. The dynamic global vegetation models (DGVMs) from TRENDY v9, and the availability of their simulations of the area and gross primary productivity (GPP) for different plant functional types (PFTs) ^a.

DGVMs	PFTs									PFT distribution	Note
	EBF	ENF	DBF	DNF	SHR	C3G	C4G	C3C	C4C		
CABLE	YES	YES	YES	YES	YES	YES	YES	NO	NO	Prescribed	
CLASS	YES	YES	YES	YES	NO	YES	YES	YES	YES	Prescribed	
CLM5	YES	YES	YES	YES	YES	YES	YES	YES	YES	Prescribed	C4C includes Corn, Millet, Sorghum and Sugarcane
JSBACH	YES	YES	YES	YES	YES	YES	YES	YES	YES	Prescribed	C3C includes C3 crop and C3 pasture; C4C includes C4 crop and pasture
JULES	YES	YES	YES	YES	YES	YES	YES	YES	YES	Simulated	C3C and C4C area the same as JSBACH
LPJ_GUESS	YES	YES	YES	YES	NO	YES	YES	YES	YES	Simulated	DBF includes boreal needleleaf summergreen, temperate broadleaf summergreen, temperate shade-intolerant broadleaf summergreen
											DBF includes tropical broad raingreen, temperate and boreal broad summergreen
LPX_Bern	YES	YES	YES	YES	NO	YES	YES	YES	YES	Simulated	C3C includes cropland C3 herbaceous and pasture C3 herbaceous; C4C includes cropland C4 herbaceous and pasture C4 herbaceous

ORCHIDEE	YES	YES	YES	YES	NO	YES	YES	YES	YES	Prescribed	DBF includes tropical broadleaf raingreen, temperate and boreal broadleaf summergreen
ORCHIDEE -CNP	YES	YES	YES	YES	NO	YES	YES	YES	YES	Prescribed	Same as ORCHIDEE, but not used, as GPP values are quite low
SDGVM	YES	YES	YES	YES	NO	YES	YES	YES	YES	Simulated	
YIBs	YES	YES	YES	NO	YES	YES	YES	YES	YES	Prescribed	

^a The PFTs include evergreen needleleaf forests (ENF), evergreen broadleaf forests (EBF), deciduous needleleaf forests (DNF), deciduous broadleaf forests (DBF), shrubs (SHR), C₃ grasses (C3G), C₄ grasses (C4G), C₃ crops (C3C), and C₄ crops (C4C).

It is a great point from the reviewer that, for models that do not have dynamic vegetation enabled, the parameters feedback on PFT distribution is not possible. We have updated the statement accordingly to reflect this point (L258). However, since our main result is about inter-model GPP difference caused by inter-model difference in PFT distribution, we are less focused on exploring the dynamic changes in PFT over time for DGVMs. Therefore, in figure S5, we are simply discussing a case if we replace PFTs (even static) in DGVMs with a consistent RS PFT, what would the estimated GPP be.

R3C4: Methods – emulators

I really like the approach of developing PFT specific emulators to test PFT uncertainties. I appreciate this is an additional analysis prompted by a suggestion in my original review. I think it's a very novel approach, so much so that there's part of me that thinks it's such a nice and novel analysis that it could be its own paper.

However, I need more information here. What were the exact inputs to the emulator? Over how many grid cells was each PFT specific emulator developed (e.g. grid cells selected across a range of climate zones for each PFT?) and if so, how were those grid cells selected? Or was it a global emulator for each PFT? What GPP data were used to train the model? Over what time period was the outputs of the random forest model? Was a spin up and transient included in the emulator? Was each emulator independently tested/validated? If so, what were the results of that? All of this feeds into how we can interpret Fig. S5.

Response: We also feel excited about emulators, as they seem to suggest we can at least use machine learning to quickly “run” DGVMs. We are happy to provide more details below:

Specifically, the inputs to each emulator included annual climate forcings (temperature, precipitation, radiation, pressure and wind speed) during 2001-2019 from CRU JRA v2.1 datasets, which was consistent with the climate forcings for DGVMs in TRENDY v9. Each emulator was constructed globally for each PFT, using pixel PFT GPP outputs derived from the respective DGVMs in each year. We excluded the LPJ-GUESS model from our emulator analysis, as it reports GPP rates in per unit of grid cell, rather than per unit of PFT area like other models. We did not explicitly include a spin-up period, as the emulator training was directly based on steady-state annual DGVM outputs. Additionally, each emulator was independently validated through cross-validation, showing consistently high predictive

accuracy (mean R^2 values > 0.8 for all PFTs, Table S4). We have now included Table S4 and methodological details in the revised manuscript (L480-493).

Table S4. The coefficient of determination (R^2) of emulators for PFT GPP simulations

		CABLE	CLASSIC	CLM5	JSBACH	JULES	LPX-Bern	ORCHID EE	SDGVM	YIBs
EBF	cal	0.98	0.99	0.87	0.99	0.98	0.38	0.93	0.99	0.95
	val	0.98	0.99	0.86	0.99	0.98	0.54	0.93	0.98	0.95
ENFt	cal	0.97	0.98	0.99	0.89	0.94	0.27	0.85	0.97	0.74
	val	0.97	0.98	0.99	0.88	0.94	0.76	0.85	0.97	0.72
ENFb	cal	0.96	0.98	0.98	0.80	0.93	0.95	0.84	0.89	0.83
	val	0.96	0.97	0.97	0.78	0.91	0.95	0.84	0.88	0.82
DBF	cal	0.97	0.99	0.97	0.99	0.98	0.46	0.91	0.95	0.77
	val	0.97	0.99	0.97	0.99	0.98	0.84	0.91	0.95	0.76
DNF	cal	0.98	0.97	0.92	0.87	0.94	0.91	0.61	0.93	0.77
	val	0.98	0.97	0.92	0.85	0.94	0.86	0.61	0.93	0.76
SHR	cal	0.98	0.98	0.95	0.99	0.97	0.44	0.84	0.96	0.66
	val	0.98	0.98	0.95	0.99	0.97	0.38	0.84	0.96	0.66
C3Gw	cal	0.96	1.00	0.96	0.98	0.99	0.77	0.78	0.90	0.70
	val	0.96	0.99	0.96	0.97	0.99	0.60	0.78	0.90	0.69
C3Gk	cal	0.96	0.98	0.91	0.94	0.98	0.44	0.84	0.96	0.59
	val	0.96	0.98	0.91	0.94	0.98	0.38	0.84	0.96	0.59
C3Gc	cal	0.97	0.99	0.91	0.96	0.96	0.86	0.65	0.92	0.80
	val	0.94	0.98	0.91	0.95	0.95	0.86	0.64	0.91	0.77
C3C	cal	0.96	0.96	0.98	0.97	0.99	0.88	0.84	0.95	0.74
	val	0.96	0.95	0.97	0.97	0.99	0.84	0.83	0.95	0.73
C4G	cal	0.97	0.99	0.91	0.96	1.00	0.93	0.91	0.84	0.77
	val	0.97	0.99	0.90	0.96	1.00	0.92	0.91	0.83	0.77
C4C	cal	0.97	0.95	0.90	0.95	0.99	0.84	0.86	0.96	0.80
	val	0.97	0.94	0.90	0.94	0.99	0.94	0.86	0.96	0.78

Note: cal and val indicate the R^2 for training and validation, respectively.

R3C5: Methods – Remote sensing estimates and uncertainties of global PFT distribution

I appreciate the consideration the authors have taken in improving the methods section by adding details and clarifications. I am still a little uncertain how the uncertainties in both the underlying LC map, and the “cross-walking” to PFTs (including uncertainties in bioclimatic limits and phenology based thresholds to separate out different climate zones) has been considered. The authors discuss this, but a) I still don’t understand some details; and b) I still don’t think all these uncertainties have been considered.

For example:

“We first obtained the uncertainties generated during mapping each land cover type by machine learning classifiers, and then assigned the uncertainties to PFTs based on translation rules from land cover to PFT maps (Bonan et al., 2002).” ♦ What are the machine learning classifiers in this case? How does that give you the full uncertainty related to the LC map? And have uncertainties related to the translation rules between land cover and PFT map in the MODIS product also been considered – I can’t see any consideration of uncertainties in the cross-walking from LC to PFT in the Sulla-Menashe? In fact at Lines 407-408 the authors say “The uncertainty in MODIS PFTs was based on the reported uncertainties in the initial land cover mapping (Sulla-Menashe et al.)”. This explicitly states that only uncertainties in the LC maps have been considered. Not any uncertainties in translating those (often broad) LC classes into PFT fractions. As we described in Hartley et al. (2017) the range of % you can give for some classes, especially in the very mixed and sparsely vegetated cover types, is almost impossible to know and very subjective. Similarly, you used additional bioclimatic limits from Bonan et al. (2002) and phenology based thresholds to further define PFTs according to their climate zone. However, these bioclimatic limits and phenology based (e.g. GDD)

thresholds have uncertainties as well, and this is also not represented in your MODIS error bars. Some of the phenology based metrics – e.g., whether you could GDD above a temperature threshold of 0 or 5°C – are very arbitrary.

Also:

“Uncertainty for the ESA CCI product was derived by digitizing reported uncertainties” ◇ what does digitizing reported uncertainties mean? So I see how the ancillary tree cover dataset can be used to estimate tree %, but what about all the other LC classes such as sparse vegetation – how did you assign fractions to each PFT? Did you use the ESA CCI cross-walking table for that? And therefore did you consider uncertainties in those very subjective cross-walking values, as described in Hartley et al., 2017.

Response: We appreciate the detailed feedback and agree that quantifying uncertainty in RS PFT distributions is a challenging task, as demonstrated by Hartley et al. (2017) and the early comment from the reviewer. Given these complexities, we simplified our approach by relying on reported uncertainties from the literature, rather than developing a framework to recalculate them, which would merit another study. Please also kindly refer to our response to R3C2.2, in which we introduce our updated method to quantify both own and inter-product uncertainty in PFT maps.

We note there is a big gap in quantifying the own uncertainty of each RS PFT map, and some maps did not provide any sources of uncertainties in their product at all (Table S3). While MODIS and ESA CCI uncertainties have been reported, existing literature did not consider all uncertainty sources or only for some PFT classes (Hartley et al., 2017; Sulla-Menashe et al., 2019). For MODIS, only the uncertainty of mapping land cover (LC) was provided, lacking the uncertainty for translating from LC to PFT. For ESA CCI, the reported uncertainty accounts for both mapping and translation, but only tailored to major vegetation classes — Tree, Shrub, Grass, and Bare. In our original version, we assigned the uncertainty from the "Tree" class to tree-specific PFTs (EBF, ENF, DBF, and DNF), from "Shrub" to Shrub PFTs, and similarly, the Grass uncertainties were uniformly applied to cold, cool, and warm grasses.

After some consideration, we think that our original approach may have underestimated RS PFT uncertainties, in agreement with the reviewers' comments. In this round of revision, we adopted five widely-used RS PFT maps and considered inter-product uncertainty to avoid over-constrained GPP resulting from insufficient consideration of uncertainty when using an individual RS PFT map. We have updated all figures and discussions using the inter-product uncertainty, please see our response to R3C2.

R3C6: Results – Relationship between PFT GPP and area

I understand perhaps better than I did in the original review what point the authors are trying to make here. Still, I would say that for some PFTs the linear relationship is not as strong (lower R² values) and the fact that the R² values are often lower when they are not forcing the regression line through zero also gives me pause. The logic for forcing the regression through zero is sound, but the authors are also just trying to see what the sensitivity of the model GPP is to the PFT area. In cases where the R² is much lower, then actually what the non zero constrained regression is telling you is that actually the total GPP is not that sensitive

to the area of the PFT as it is for other PFTs. This is especially the case for C3 grasses which are one of the most abundant PFTs worldwide (if you add up the area of the individual C3 categories in this study). So when the authors say in their response to R1C2 that “it is very surprising to see ALL models show EBF has similar productivity per area (i.e., the slopes of the total GPP – area relationships for each PFT in Fig. 1)” I would say I agree, but I also think that’s not the case for all PFTs.

I think this analysis is interesting though because the differences across PFTs in terms of the strengths of the linear regression (unconstrained R² values) provide insights into which PFTs have very similar representation across different models (and therefore where the prescribed PFT area is causing most of the difference in total area) versus the PFTs where area plays a role, but clearly other aspects of modeling that PFT are contributing to the spread in total GPP (which then gives lower R² values).

Some of the PFTs that clearly have a strong linear relationship between total GPP and area are not that surprising – i.e. C4 grasses have had little attention paid to them compared to C3 grasses, and so it is perhaps predictable that all models simulate a similar GPP per area as their underlying representation is very similar. In contrast, more work has been done on C3 photosynthesis – for example, different models for calculating C3 photosynthesis, different parameter values (some of which have been calibrated against more numerous datasets for C3 grasses), including optimality-based approaches, and including temperature acclimation, etc – and therefore I assume there are differences across models, and that explains the lower R² values. (I also note that Figure S1 is not cited in the main text). Deciduous needleleaved forests are another example that have had relatively little attention paid to them compared to other PFTs and therefore the R² values are quite high.

Response: Thank you for the comments, which further expanded the comment R3C2.1. We have therefore added a paragraph to specifically discuss the difference in strength of the relationships between PFTs, in L303-307.

“This framework also helped explain the variability in the strength of GPP-area relationships among PFTs (i.e., R² and slope values in Fig. 1) – the relatively lower R² values, such as those for C₃ grasses and crops, may indicate greater inter-model difference in LAI or leaf photosynthetic rates compared to other PFTs.”

R3C7: Results – Updated land carbon cycle estimation using remotely sensed PFT area
Given the issues I mentioned above I am afraid I do not trust the narrow range of uncertainties used for either MODIS or ESA CCI. See my comments in the general section on suggestions for addressing the key caveats above.

Fig. 2: Caption include the fact that the DGVMs are from TRENDY and should say which version, even though this is in the methods (as many people won’t read the methods). In fact I would make it clear what is going into each of these estimates in the figure caption. Which products are going into MLEC for example, just FLUXCOM or a range of different upscaled EC products?

Response: Please kindly refer to our answer to R3C2.2 about the uncertainty of RS PFT maps and associated reduction in GPP estimates.

We have also clarified the version of TRENDY DGVMs (v9), as well as the version names for ML_{EC} products in both figure captions and Method section. Specifically, ML_{EC} includes the RS_METEO output from FLUXCOM – ensemble estimates upscaled using different machine learning methods and meteorological forcings (Jung et al., 2020), as well as X-BASE product from the FLUXCOM-X (Nelson et al., 2024).

R3C8: Results – Contribution of PFT change to global GPP change

This section is really nice and interesting and yet it is a little lost in the overall message about constrained PFT area leading to constrained GPP based on the strength of the linear relationship between PFT GPP and area. I would have liked to see more here, although I appreciate there's not much space in a letters journal. What is very curious about this is that we expect to see C4 grass outcompete at higher CO₂ concentrations, if I'm remembering correctly. This is not directly discussed although the authors mention the potential water savings under elevated CO₂. I would suggest the authors expand this though to suggest future studies check that that is indeed what is happening (water savings) and the increased C3 is not because of incorrect process representation in the models.

Response: Thank you. We have expanded the discussion (L189-190) to suggest that “the expansion of C₄ was likely due to its anatomical traits, which confer higher water-use efficiency”, supported by theoretical understanding (Way et al., 2014) as well as global isotope observations (Eggels et al., 2021). The C₃ and C₄ WUE are very different in theory, and C₄ is expected to be more WUE. If models use conventional Farquhar approach for C₃ and Collatz approach for C₄, they all should be able to get higher WUE from C₄ than C₃, though the magnitude may vary depending on model parameterization.

R3C9: Discussion

Lines 247-249: “To improve biogeography process in DGVMs, we suggest prioritizing efforts to clear hotspots of biogeography uncertainty (e.g., those identified in Fig. 3) and to develop a more reliable global PFT map” ◇ Yes absolutely! Actually most of these models have no dynamic vegetation nor vegetation demography that is causing any kind of PFT change. They all have prescribed maps. So yes, a more reliable global PFT map is needed for models that do that! And yet, as the authors have shown, the two most widely available maps have very different areas, which, as the authors have also shown, feeds into differences in total GPP (though I just wouldn't place as much emphasis on the narrowing of uncertainty in that total GPP estimate as the authors do). This is the key message of the paper here I think, but I believe it can be framed more clearly.

Response: Following the suggestion from the reviewer, we have avoided putting too much emphasis on narrowing of the GPP uncertainty. Instead, we updated Fig. 3 using five widely-used RS PFT maps and highlighted the uncertainty hotspots in PFT maps. We have rephrased the statements to

“Our study provided a framework to quantify the magnitude of land carbon cycle using RS PFT maps, however, we must emphasize the considerable difference among RS PFT maps (Fig. 1; Fig. 3). A long line of research has indicated large disagreement between RS PFT maps, and those differences are still challenging to reconcile. The key difference could be relevant to the fraction of specific PFT types in the world (e.g., especially the EBF and C₃ crops). For example, MODIS suggests the EBF area is $21.56 \pm 1.23 \times 10^6 \text{ km}^2$, but ESA CCI suggests it is only $14.23 \pm 2.85 \times 10^6 \text{ km}^2$. The discrepancy in RS PFTs can propagate into the global GPP estimates, leading to an approximately 20 Pg C difference (Fig. S2).”

We have also included a paragraph to discuss the uncertainty hotspot of RS PFT maps (L212-218).

“Moreover, we found that RS PFT maps also show large differences in their estimates of PFT total area and spatial distributions (Fig. 3; Table S3). For example, there were large spreads in the estimates of global EBF area (i.e., from 13.97 to $21.56 \times 10^6 \text{ km}^2$) and C₃C area (from 7.66 to $16.34 \times 10^6 \text{ km}^2$) across the five remote sensing products. The hotspots of biogeography disagreement among RS PFT maps include EBF in the eastern U.S. and western Europe, ENF in northern Canada and eastern Russia, DNF in North America and Europe, and SHR and C₃G across the globe.”

R3C10: Line 249-250: I guess I would say be careful with the use of “alternative” in this statement: “Using a consistent biogeography baseline, such as a PFT map from remote sensing as a diagnostic alternative” because these PFT maps are being used in models. I know ORCHIDEE and SDGVM use the same ESA CCI map (which also makes me wonder about the circularity in this approach – I would like the authors to comment on that).

Response: We have rephrased the statement to “Using a consistent biogeography baseline in DGVMs can be a first step to reduce PFT-induced uncertainty in land carbon cycle simulations.”

Regarding the issue of circularity arising from the use of the ESA CCI map, we would like to clarify that although some DGVMs use the ESA CCI map, the implementation of this map varies largely among models:

1) We note clear differences in the use of ESA CCI product in ORCHIDEE and SDGVM. ORCHIDEE distinguishes among 15 PFTs, whereas SDGVM differentiates only 10. For instance, ORCHIDEE divides C₃ grasses into temperate, tropical, and boreal categories, while SDGVM aggregates all C₃ grasses into a single category. These differences imply distinct PFT-specific parameterizations and structural frameworks, potentially resulting in different GPP rates per unit area.

2) Even when models employ ESA CCI maps, they often integrate with additional land-cover datasets. SDGVM, for example, merges ESA CCI 2014 land-cover data (Poulter et al., 2015) with LUH2v2h cropland and pasture datasets.

3) In addition, SDGVM enabled a dynamic vegetation component, while ORCHIDEE used a prescribed static PFT map. The two models also simulated on different spatial resolution – SDGVM at 0.5° and ORCHIDEE at 0.25°.

Given the varying approaches for implementation of ESA CCI data across DGVMs, the PFT distributions in these models are not likely to be the same with each other (also shown in our Fig. 1). Additionally, in this round of revision, we have updated the main result based on five RS PFT maps, rather than a single one, to further avoid the issue of overly relying on one RS PFT map.

R3C11: Lines 256-261: This is only true for dynamic vegetation models, of which there are only 4 in the TRENDY v9 ensemble. Otherwise, photosynthesis rates etc do not feedback on PFT area. So I would revise this section accordingly.

Response: Please kindly see our response to R3C3.

R3C12: Lines 269- : I like the additional emulator analysis and I really appreciate the authors attempting to address this comment from the first round of review. I have one minor comment and one concern:

My concern is that I don't think the larger spread (smaller reduction in uncertainty in EM(DGVMs,MODIS)) is because the models have tuned their structure to the wrong location. Or at least, I can't see how that would be the case. Let's say C4 grass is in the completely wrong location in many models because the bioclimatic limit used to separate out C3 and C4 grasses is wrong. If so, what does that imply, that their V_{cmax} parameter for example is adjusted to the climate of the wrong location to match in situ flux tower data? That seems convoluted to me, although I think maybe I see the argument. I'd be willing to bet though that most PFTs are not in the completely wrong location (mostly based on prescribed PFT maps). It would mean major biases in parameters or model structure to account for that. What I think is maybe more likely, or explains more of the difference between EM(DGVMs,MODIS) and DGVM_MOD is what I've said throughout this review – that the uncertainty reduction from constraining with MODIS (or ESA CCI) PFT map area is too high. I would guess that if you had more realistic uncertainties on your PFT maps, your constrained GPP would also be about 40-50% as in EM(DGVMs,MODIS). It still shows a big importance of PFT map, but again, I think it just cannot explain the majority of the spread across models. Still, this nice emulator result shows that differences simply due to the PFT map are a greater issue than anyone would bet.

My minor comment is that I am not sure I would describe it as a difference between biogeography and biogeochemistry (Responses to R1C7 and R2C3 and new text describing Fig. S5) (with my issue being mostly with the use of biogeochemistry). If the location of the PFTs is wrong (e.g., Bias type 1 in Response to R2C3), then that's also a biogeography issue right? Or am I misunderstanding? If the location is wrong then it could be that the bioclimatic envelopes for example are not right, and I would call that a biogeography "process" or limitation (that is intimately tied to area). On the other hand, you may have an issue with temperature optima, which would prohibit certain PFTs from growing in an area. I would call that a biogeochemistry (or plant physiology actually) process (could also be related to

phenology thresholds or ecohydrology if not enough water was available). So it's just a question of wording but I perhaps would more simply say that what you said in response to R2C3 – that there are two types of biases related to area and location, and that biases related to area only were accounted for by the analysis in Figure 1. But Figure 3 showed the second bias related to location distribution, and so you have done this analysis. Basically, I think the use of biogeochemistry here is confusing and also misleading given that for most models any “biogeochemistry/plant physiology” etc process is not feeding back on the distribution anyway. If the distribution is wrong for most models then it's because the prescribed PFT map is wrong (except for the dynamic vegetation models).

Response: We appreciate the discussion from the reviewers on the interpretation of our Fig. S5 (the emulator analysis, now Fig. S7), and how much the figure tells us about the relative importance of biogeography and biogeochemistry in the uncertainty of modelling GPP.

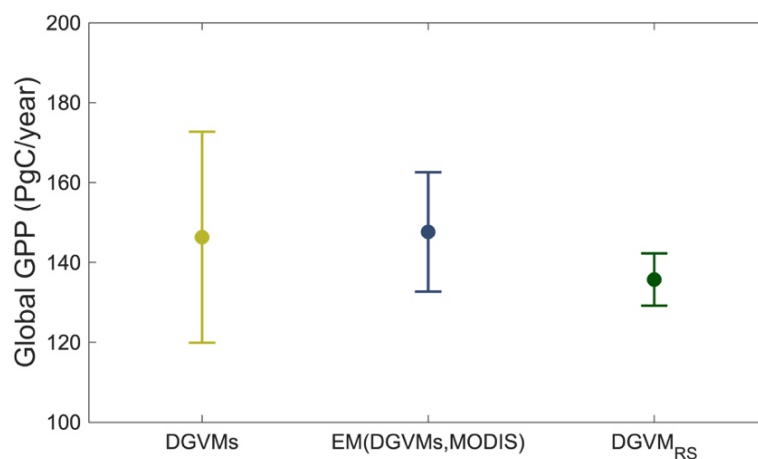


Figure S7. The spread of global GPP estimates from original DGVMs and DGVMs forced by MODIS PFT map. The three group of models are 1) DGVMs – the original group of DGVMs, with biased PFT area and location; 2) EM(DGVMs, MODIS) – the emulators of DGVMs forced with MODIS PFT maps, meaning the PFT area and locations are both “corrected”; 3) DGVM_{RS} – the constrained DGVMs simulations, meaning only the PFT area was “corrected” but location remained biased.

Perhaps it is good for us to establish some baseline understanding first. We quote reviewer’s comment below.

1. “I’d be willing to bet though that most PFTs are not in the completely wrong location (mostly based on prescribed PFT maps).” – this is a complex case. Below we show the maps of the disagreement of PFT maps across DGVMs (Fig. S6), and they can be really different and thus wrong (at least for some)! For example, a few models predict abundant EBF distributions (> 10%) in parts of Siberia and ENF in tropical South America and Africa.

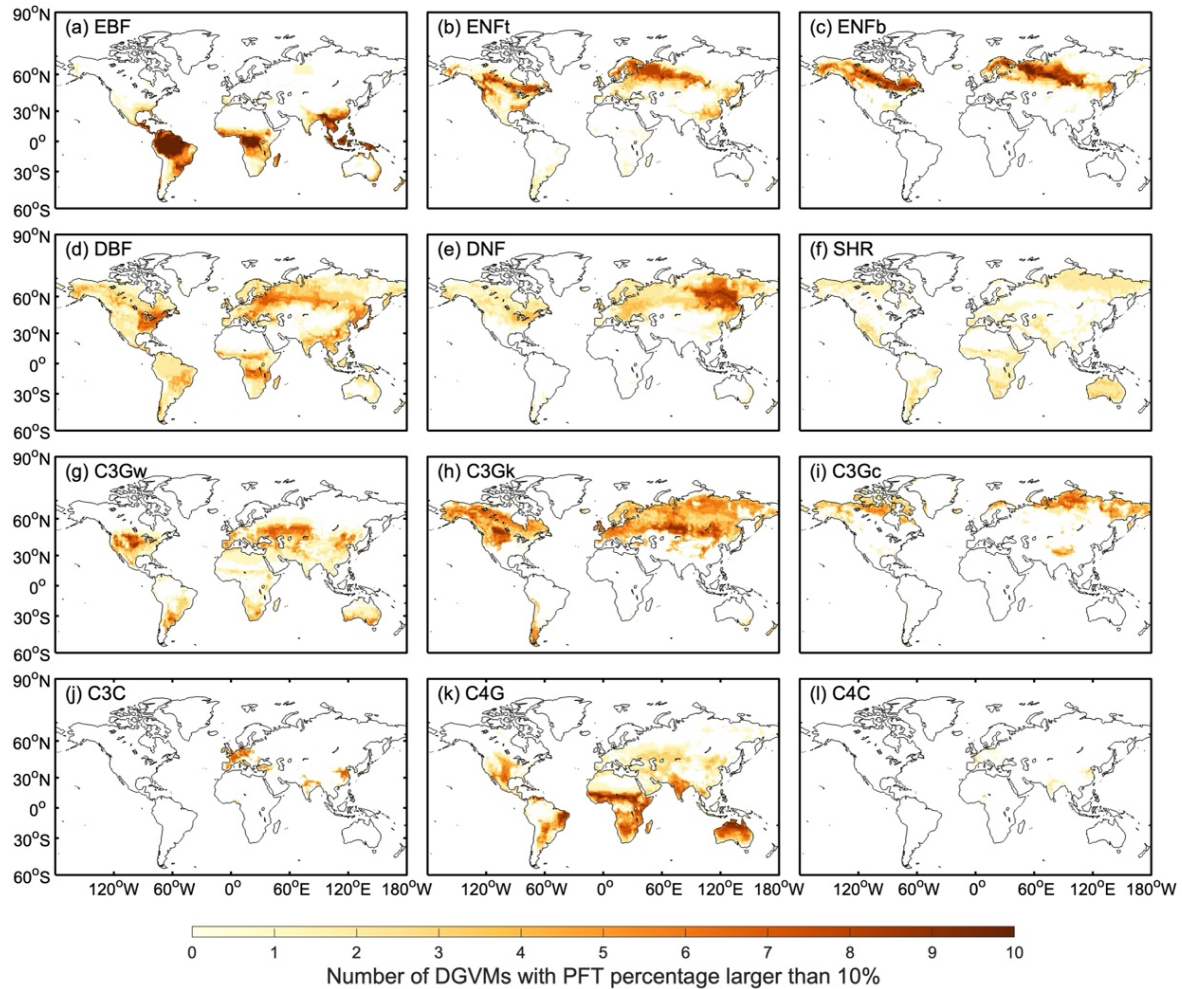


Figure S6. The difference in the distribution of PFTs in DGVMs, quantified by another method. Here we presented the number of DGVMs in each grid cell for each PFT. The number of DGVMs where PFT percentage larger than 10% in the same grid, with higher values indicating a greater consistency. The abbreviations for PFTs correspond to those in Figure 1.

2. “that the uncertainty reduction from constraining with MODIS (or ESA CCI) PFT map area is too high”. Originally, we reported 86% reduction, and now it is 75% after we used five RS PFT maps to constrain GPP. Even with this updated result, biogeography still emerges as a main factor controlling the uncertainty of the land carbon cycle modelling.

3. Thus, the intriguing finding illustrated by Fig. S5 (now Fig. S7) remains valid. When we correct area ($DGVM_{RS}$), uncertainty in GPP simulations reduced by 75%; when we correct both area and location ($EM(DGVMs, MODIS)$), uncertainty reduced by 46%. It means that, if we place PFT at the right location, it may not immediately reduce the uncertainty, because the new parameters coming with the replaced PFT may actually inflate the uncertainty. We do feel without a proper MIP and scenario run, we are unable to provide a clear picture, but our (best) guess is that the parameters of the correct PFT for that location might have some issues.

In this analysis, since we are discussing *inter-model* differences using multi-year average, the issue whether parameters feedback on PFT distribution is not a direct concern, given their

static PFT maps incurred a large difference in GPP estimates already. Please also see our reply to R3C3.

R3C13: Reply to response to R1C3 in response to reviewers: I would perhaps disagree that PFTs can be directly linked to spectral signatures. In fact, it is very hard to distinguish trees from grasses, even in hyperspectral data and even when you add LiDAR data (as my PhD student Rubaya Pervin found in drylands given there's also a strong spectral signature from dry grasses and/or biocrusts). In the past this has prompted the suggestion of using "optical functional types" (Ustin and Gamon, 2010) where global vegetation would be classified based on the clustering of spectral signatures. To my knowledge, no one has really tried this at regional or global scale. I would bet this would look different to the PFT maps we currently use and would be hard to match with anything we represent in models. So the question remains, how can we use satellite data to infer plant functional type fractional cover? The fusion of tree cover fraction in the ESA CCI product is a great advance. But it also depends on what is counted as a tree (height thresholds etc).

Response: Thank you for sharing your thoughts on remote sensing-based PFT mapping. We apologize for the somewhat broad statement on "spectral signature", and we have updated it to "spectral signature and its variation" (such as phenology, snow cover) to reflect what is really being implemented in many RS PFT mapping work.

We encourage community input to improve PFT mapping, though addressing this grand challenge is beyond the scope of our study here – we intend to explore the global vegetation biogeography in the future, not only based on remote sensing, but also based on ecological theories and the insights they might provide to PFT distributions.

R3C14: Reply to response to R1C6 in response to reviewers: Yes this is very interesting to see the range of estimates, and also the reporting of more recent studies that all have higher estimates. In fact, the higher mean value of GPP as constrained with MODIS is one of the most interesting things about this paper (although this is somewhat dampened by the fact that ESA CCI maintains the typical value of around 120PgC), although other methods have a higher value as the authors show in Fig. 2a. As I am signing this review I apologise for this shameless self-citation, but a paper in which I constrained GPP using SIF data (one of the earlier, simpler approach using just one model) also found a posterior (post parameter optimization) global value of just over 160PgC (MacBean et al., 2018). Please do not feel you need to include this as a reference, it is more for your interest. Since I wrote that paper in the back of my mind has been the thought that yes ok we constrained the high positive bias in ORCHIDEE (pre optimization it was above 190PgC), but we have not constrained it enough. I hadn't quite processed until I read the authors response to the original review that there are indeed multiple studies pushing at this higher value (and hadn't seen the Liu paper). As the authors say, the value of 120-130PgC has been talked about for Beer's 2010 Science paper if not before and is the number that is most often quoted. It has been 130PgC in the the GCB for at least the last two years. So I think it's great that the authors have moved away from a "correct" magnitude and instead show the uncertainties. Still, when you put all these studies

together with the new study presented here, it does feel like the evidence is mounting for a higher value.

Response: We thank the reviewer for sharing her earlier study, and we are glad to see another piece of evidence supporting a higher GPP value. We cited it to further support our argument that based on the evidence so far, it is likely global GPP is greater than we thought before. The jewel is out there again. Our study may motivate the community to think about why we should have a higher GPP than we thought before.

Minor points:

R3C15: Line 29: would suggest “biogeography-climate” instead of “biogeography – climate” interactions. First time I read it I thought the words after the hyphen were making a different point.

Response: Thanks. We have revised this accordingly.

R3C16: Lines 54-56: another limitation for MODIS GPP (or any GPP derived from the production efficiency/LUE method) and the machine learning approaches I believe is the lack of accounting for the effect of photosynthesis on LUE (De Kauwe et al., 2016).

Response: Yes, general LUE-based models, such as the one used in MODIS GPP product, do not account for elevated CO₂ (De Kauwe et al., 2016) and their considerations of climate impacts are quite empirical (Medlyn, 2011). We have now added this point in L59.

R3C17: Line 71: would maybe not use the word “prescribe” here (to be pedantic) – maybe just that in process-based models these processes are specific to each PFT ?

Response: We have removed the word ‘prescribe’ and updated the sentence to “the current framework for improving land carbon cycle simulations in DGVMs implies the importance of biogeography, given that changes in model parameters, structures, and even the emerging climate sensitivities of simulated carbon fluxes can also be driven by PFT change, as DGVMs largely regard these processes as PFT-specific”.

R3C18: I cannot see CLM5 in Fig. 1j? I’m guessing that’s why you have axes upper limits that are much higher than the main cluster of models

Response: CLM5 does not appear in Fig. 1j because, although the TRENDY v9 dataset indicates that CLM5 provides C₃ crop data, upon careful examination of the output files we found the data for C₃ crop to be entirely NaN across all pixels. Therefore, we were unable to include CLM5 C₃ crop in this analysis.

R3C19: I like the new additional analysis presented in Fig. S6. Nice!

Response: We are pleased to hear that the additional analysis is helpful.

R3C20: Data Availability Statement needs to include where you got the TRENDY v9 GPP and PFT from as it’s actually not on the global carbon budget page that links to Mike O’Sullivan’s Github (<https://mdosullivan.github.io/GCB/>). Also, would you consider providing the emulators in a publicly available repository for people to test other PFT maps?

Response: We obtained the TRENDY v9 data from the project coordinator, Dr. Stephen Sitch and have now revised this statement to,
“DGVM simulations from TRENDY v9 were obtained from Dr. S. Sitch (s.a.sitch@exeter.ac.uk) and are now available at <https://mdosullivan.github.io/GCB/>.”

Moreover, we have now made the emulator code and associated models publicly available via the Zenodo repository at <https://doi.org/10.5281/zenodo.15752958>.

References

De Kauwe, M. G., Keenan, T. F., Medlyn, B. E., Prentice, I. C. and Terrer, C. (2016) Satellite based estimates underestimate the effect of CO₂ fertilisation on net primary productivity. *Nature Climate Change*, 6, 892-893.

MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., ... & Disney, M. (2018). Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. *Scientific reports*, 8(1), 1973.

Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, 186(4), 795-816.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response: We greatly appreciate your review and have endeavoured to use all comments to improve our study. Please kindly find our detailed point-by-point responses to reviewers 3 and 4 altogether above.

References:

- Buchhorn, M. et al., 2020. Copernicus global land service: land cover 100m: version 3 globe 2015-2019: algorithm theoretical basis document. Zenodo.
- Dantas De Paula, M. et al., 2025. Including the phosphorus cycle into the LPJ-GUESS dynamic global vegetation model (v4.1, r10994) – global patterns and temporal trends of N and P primary production limitation. *Geoscientific Model Development*, 18(7): 2249-2274.
- De Kauwe, M.G., Keenan, T.F., Medlyn, B.E., Prentice, I.C. and Terrer, C., 2016. Satellite based estimates underestimate the effect of CO₂ fertilization on net primary productivity. *Nature Climate Change*, 6(10): 892-893.
- Eggels, S., Blankenagel, S., Schön, C. and Avramova, V., 2021. The carbon isotopic signature of C₄ crops and its applicability in breeding for climate resilience. *Theoretical and Applied Genetics*, 134(6): 1663-1675.
- Fatichi, S., Pappas, C., Zscheischler, J. and Leuzinger, S., 2019. Modelling carbon sources and sinks in terrestrial vegetation. *New Phytologist*, 221(2): 652-668.
- Friedlingstein, P. et al., 2023. Global Carbon Budget 2023. *Earth System Science Data*, 15(12): 5301-5369.
- Harper, K.L. et al., 2023. A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models. *Earth System Science Data*, 15(3): 1465-1499.
- Hartley, A.J., MacBean, N., Georgievski, G. and Bontemps, S., 2017. Uncertainty in plant functional type distributions and its impact on land surface models. *Remote Sensing of Environment*, 203: 71-89.
- Hu, Z. et al., 2018. Joint structural and physiological control on the interannual variation in productivity in a temperate grassland: A data-model comparison. *Global Change Biology*, 24(7): 2965-2979.
- Jung, M. et al., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, 17(5): 1343-1365.
- Keenan, T.F. et al., 2023. A constraint on historic growth in global photosynthesis due to rising CO₂. *Nature Climate Change*, 13(12): 1376-1381.
- Liu, H. et al., 2020. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth Syst. Sci. Data*, 12(2): 1217-1243.
- Medlyn, B.E., 2011. Comment on “Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009”. *Science*, 333(6046): 1093-1093.
- Nelson, J.A. et al., 2024. X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X. *Biogeosciences*, 21(22): 5079-5115.
- Poulter, B. et al., 2015. Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative. *Geoscientific Model Development*, 8(7): 2315-2328.
- Prentice, I.C. et al., 1992. Special Paper: A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate. *Journal of Biogeography*, 19(2): 117.
- Rogers, A. et al., 2017. A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytologist*, 213(1): 22-42.
- Sulla-Menashe, D. and Friedl, M.A., 2018. User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. *Usgs: Reston, Va, Usa*: 1-18.

- Sulla-Menashe, D., Gray, J.M., Abercrombie, S.P. and Friedl, M.A., 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, 222: 183-194.
- Wang, X. et al., 2021. Beneficial effects of nitrogen deposition on carbon and nitrogen accumulation in grasses over other species in Inner Mongolian grasslands. *Global Ecology and Conservation*, 26: e01507.
- Way, D.A., Katul, G.G., Manzoni, S. and Vico, G., 2014. Increasing water use efficiency along the C3 to C4 evolutionary pathway: a stomatal optimization perspective. *Journal of Experimental Botany*, 65(13): 3683-3693.
- Zhang, X. et al., 2024. GLC_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method. *Earth Syst. Sci. Data*, 16(3): 1353-1381.
- Zhu, J. et al., 2025. Changing patterns of global nitrogen deposition driven by socio-economic development. *Nature Communications*, 16(1): 46.

Response to Reviewers

Reviewer #1 (Remarks to the Author)

R1C1: I am highly impressive by the revision. The revised manuscript has substantially addressed my concerns in term of uncertainty of vegetation biome areas by using more data products and also potential impacts of land usage change. The writing has been also substantially improved. Here I only have one minor suggestion for Dr.Zhao et al to consider. As commented in the first round of review, it would be useful to clarify some caution (i.e., by needs of incorporating underlying mechanism in future studies) of this data-based approach to reduce uncertainty of projected carbon cycling - GPP and ER here. Vegetation biogeography is emergent property in which plant individuals grow up and compete for the resources. With this, it could have uncertainty based on the transient snapshot of relationship between vegetation biome area and GPP & ER at community scales for projecting future GPP and ER dynamics.

Response: We are very pleased to learn that we have addressed the reviewer's comments, and that the manuscript has been substantially improved. Following the last minor comment from the reviewer, we agree that there is a need to exercise additional caution in future projections if the biogeography – GPP/ER relationships change in the future (L329-336).

“Under our statistical framework, the future changes in carbon fluxes are driven by biogeographic changes (i.e., PFT area) and the changes in biogeography–carbon cycle relationships (e.g., PFT area-GPP relationships). While our results emphasize the role of PFT area in modulating future GPP, we acknowledge that the changes in the slopes of PFT area-GPP relationships might also cause future GPP changes. This is likely because the slopes essentially reflect impacts of PFT LAI, leaf photosynthetic rate and other biogeochemical processes which are known to change with climate and elevated CO₂ (Piao et al., 2020; Walker et al., 2021)”.

Reviewer #3 (Remarks to the Author)

R3C1: I am satisfied the authors have addressed most of my concerns/comments and I thank them for taking the time for that. I especially appreciate their consideration of my point about RS product uncertainties and the fact they now use the spread across 5 RS products. However, I still have concerns about the use of a simple linear regression to highlight the main point of this study, which I think mirrors some of Reviewer 2's concerns, though I don't share Reviewer 2's concern that it's wrong to have GPP as the dependent variable and PFT area as the independent variable.

Response: We are glad to learn that we have successfully addressed most of the concerns from the reviewer and appreciate their insights on reviewer #2's comments. We understand that the reviewer is particularly interested in further corroborating the robustness of our linear regression approach to support our conclusion – to do so, we have followed their suggestions to use multivariate regression and linear regression not forced through zero in our analysis (Please kindly refer to our detailed responses below). These added analyses consistently support our conclusion that biogeography is a main source of uncertainty in modelling global land carbon cycle. Please kindly see our point-by-point responses below.

R3C2: I carefully considered the comment from Reviewer 2 about the scaling of GPP with area. I do not see their problem with having total GPP dependent on PFT area, statistically speaking, given GPP is the dependent variable and the PFT area the independent variable, unless I'm missing something here. However, having considered this more, I do perhaps see some of Reviewer 2's issue with using the linear regression in this way. I've tried to articulate some my concerns below.

First, I think one thing that Reviewer 2 is pointing out is that it is obvious that GPP scales with area. I appreciate (and agree with) the authors response that although it seems obvious given the equation for global total GPP, PFT area is not that discussed as a source of uncertainty in global GPP in the modeling community, and instead studies focus on differences in the way the model represents the photosynthesis, phenology etc (biogeochemistry as the authors discuss). But unfortunately I still think some readers will say it's obvious given the dependence on area. Further, let's say some models have extremely different area, so the range in PFT area is high, then the linear relationship will become more pronounced despite variance around the regression line due to differences in model structure/parameters etc (biogeochemistry) in estimating GPP per unit area.

Response: We thank the reviewer for further clarifying reviewer #2's comments. Like the reviewers, we agree there is indeed (and should be) an implicit understanding from most modelers that GPP scales with Area. What surprised and motivated us is the lack of quantification of such scaling, and the lack of understanding on the importance of this scaling compared to other model processes/parameters. Our study provides critical evidence to the community that brings out the implicit understanding - it clearly demonstrates that correct

biogeography is almost as important as, if not more important than, correct biogeochemistry in modelling global land carbon cycle.

From a statistical perspective, we fully agree with the reviewer that the strength of the GPP – Area relationships could be strong if the variance of Area is large across models. This again echoes our motivation above – there have been limited studies quantifying such variance in PFT area across DGVMs. Our study provides direct evidence that this variance in Area between models is substantial (Fig. S5 and S6), which has contributed to much of inter-model spread in global GPP simulations (Fig. 1).

R3C3: I was somewhat underwhelmed with the author's response to R3C6, and again I think this speaks to my, and perhaps Reviewer 2's, sense that the emphasis placed on the results in this study are more nuanced than the strong statements in lines ~92-96 based on a simple linear regression. In R3C6 in my last review I questioned forcing the linear regressions through zero, and the fact that for some PFTs the R² values are much lower if you don't force through zero. This wasn't addressed in the authors reply to my comment in the last round of review. Again, while it is true in reality that with zero PFT area then GPP will be zero, here the authors are trying to understand model behavior, and therefore I'm not sure these linear relationships should be forced through zero. It's hard to tell for sure, but in most cases where the R² values drop when the regression is not forced through zero it looks like the slope of the regression would be lower. That would actually give you a narrower range for the GPP values (higher reduction in GPP range), but it's hard to know where that slope would lie (what the y-intercept would be) and therefore what the actual GPP range would be. It could be quite different I think. So for full transparency I think the authors should not only report the R² values of the regressions not forced through zero but also the actual regression line (with slope and y-intercept) and therefore how the GPP range would be different with a regression line not forced through zero. And then see if the key points of this paper – the reduction in uncertainty on GPP (and ER) and the constrained range of GPP values – changes as a result of not forcing through zero.

Response: We appreciate the reviewer's suggestion to include a full analysis based on linear regression not forced through zero. Following the suggestion, we have now provided the analysis with full equations and associated statistics in our results (Fig. S10).

The analysis shows that this alternative approach constrains the global GPP spread by approximately 48% (i.e., from 26.39 to 13.69 PgC/year; Fig. S11). For global ER, we obtained a similar range of reduction in uncertainty (from 25.99 to 13.71 PgC/year). Though in theory we consider it is reasonable to force GPP-Area relationship through zero, since there will be no GPP if area is zero, we have now highlighted those new analyses in our results to give readers more information for assessments. We have also slightly rephrased L92-96 into a more conservative tone as the reviewer suggested. It now reads "The strong relationships indicate

that the inter-model discrepancy of total GPP of each PFT is *highly relevant* to the difference in the prescribed or simulated PFT area across DGVMs”

We highlighted the new analysis in L93-95: “Using alternative approaches (i.e., linear regression not forced through zero or multivariate regression), we consistently observed the strong dependence of total GPP on area for most of the PFTs (Fig. 1; Fig. S10; Table S5).”

L158-160: “Using linear regression not forced through zero, we still achieved a substantial yet smaller reduction (i.e., 48% compared to 75%) in the spread of inter-model GPP and ER (Fig. S11).”

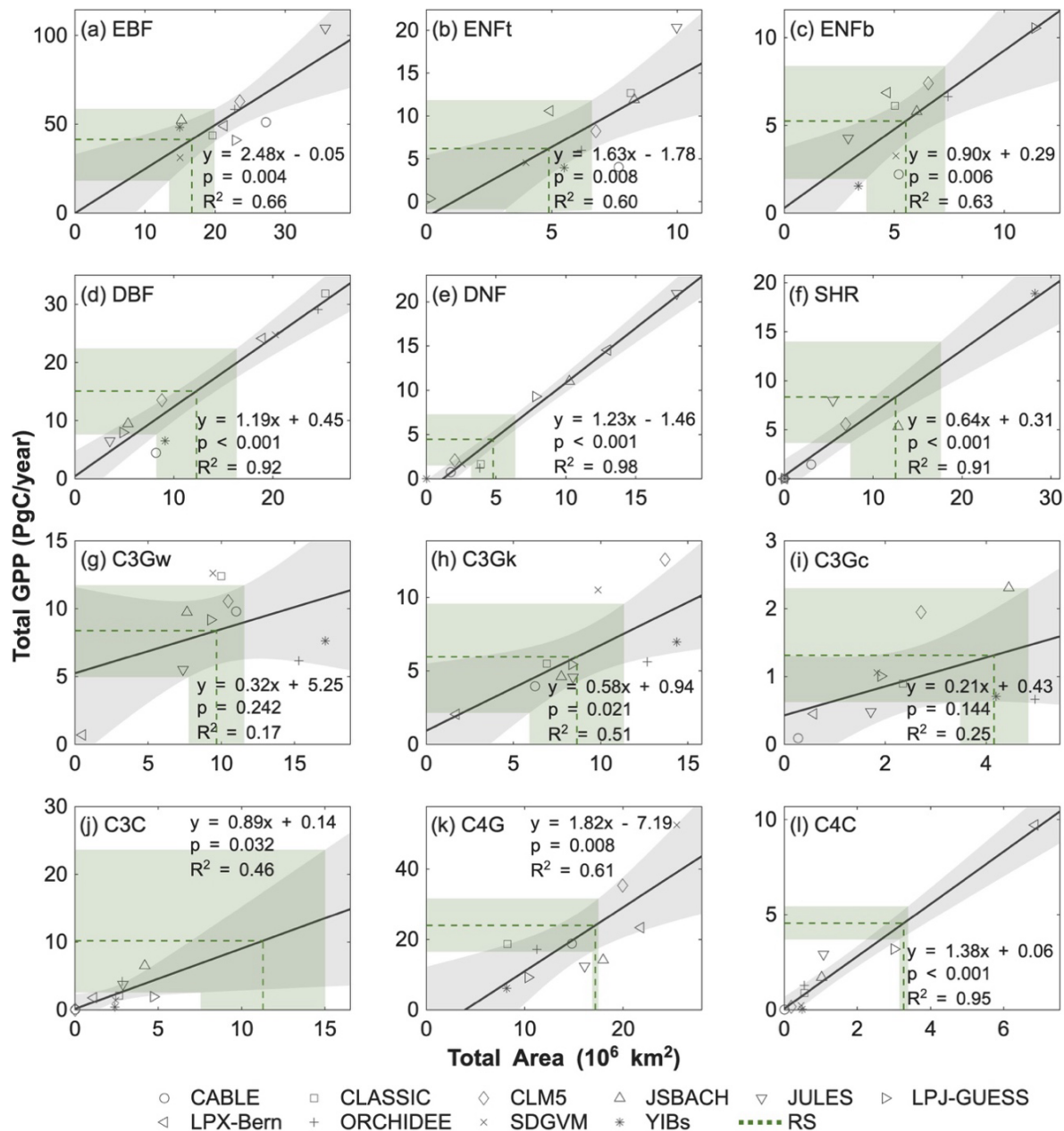


Figure S10. The relationships between total gross primary productivity (GPP) and total area for each PFT. Here we present the result similar to that in Figure 1, except that the linear regressions were not forced through zero.

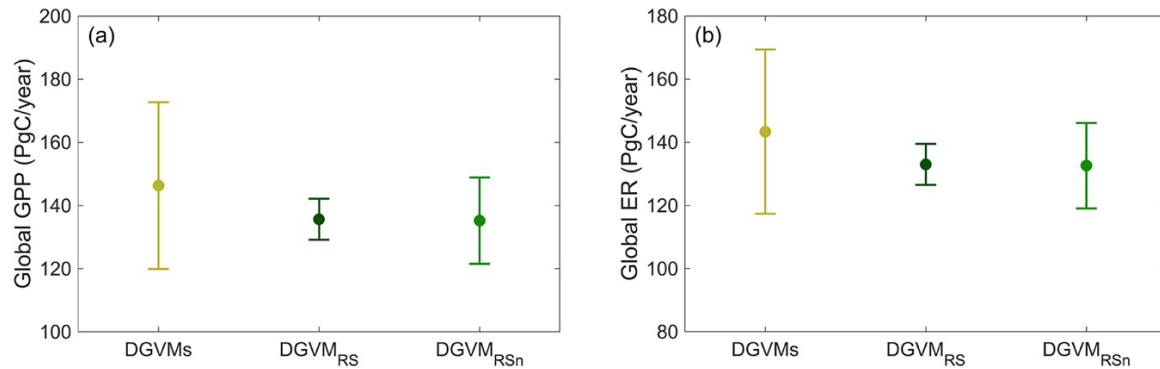


Figure S11. The constrained global gross primary productivity (GPP) and ecosystem respiration (ER). Here we present the result similar to that in Figure 2. Shown are the GPP and ER spread across models (DGVMs), constrained GPP and ER using regressions forced through zero (DGVM_{RS}), and constrained GPP and ER using regressions not forced through zero (DGVM_{RSn}).

R3C4: In response to my point in R3C6 about the fact that some PFTs do have lower R² values (and therefore again the strong statements in lines 92-96 seem a little overstated) the authors simply added “the relatively lower R² values, such as those for C₃ grasses and crops, may indicate greater inter-model difference in LAI or leaf photosynthetic rates compared to other PFTs.” First, I wouldn’t use “may” here. If there’s spread around that line it must be that there’s an inter-model difference in how the models are simulating other components of total GPP such as phenology and photosynthesis, plant water uptake, nutrient dynamics etc (factors that contribute to A_{leaf} and LAI). But second, I’d appreciate seeing a more nuanced discussion about the fact that lower R² values for some PFTs mean for those PFTs there is more disagreement among models in the representation of the biogeochemistry/physiology etc. The point is that where we don’t have that much agreement among models, PFT area will become a much stronger predictor of total GPP because the models don’t differ that much in their parameterization of the biogeochemistry (perhaps for reasons I gave in the last round of revision, such as PFTs that haven’t had as much focus on how we model them, such as C₄ grasses). Again, a reader may think this is obvious and so I urge the authors to discuss this in a more nuanced way.

Response: We first followed the reviewer’s suggestion to rephrase L92-96 (please see our response to R3C3) and remove the word “may” in L317.

We secondly tried to put together a more nuanced discussion (L316-327) on the weak GPP – area relationships for certain PFT (mostly C₃ grasslands and crops).

“– the relatively lower R^2 values, such as those for C_3 grasses and crops, indicate potentially greater inter-model difference in LAI or leaf photosynthetic rates for C_3 grasses and crops compared to other PFTs. Previous studies indeed reported a slightly larger inter-model spread in the maximum carboxylation rate of photosynthesis for C_3 grasses than those for forest types (Rogers, 2014) and a likely greater inter-model LAI discrepancy for grasses (Anav et al 2013; Song et al. 2021). Consistently, our statistical analysis shows that LAI differences in C_3 grasses contribute strongly to inter-model GPP variation (Table S5), implying a considerable difference in grass LAI among models. However, without a comprehensive comparison of model protocols and structures, it remains difficult to pinpoint the specific biogeochemical processes or parameterizations responsible for the varying strength of the GPP–area relationship across PFTs.’

We must emphasize that without a proper test (which is a bit beyond our capacity), it is challenging to pinpoint the exact parameters and processes for the weak relationships. This would require comprehensive examination of model protocols and the structures. For example, the TRENDY protocol often allows flexibility (e.g., JULES can use Farquhar or Collatz equations for C_3), and the exact configuration submitted for each model is not always reported. It is also not transparent on the versions of each model applied and technical documentation of specific model runs for TRENDY project. We have therefore refrained from making too many speculations in the discussion, but we do provide clear evidence that which PFTs we should pay more attention to in future studies on biogeochemistry (e.g., C_3 grasses and crops).

R3C5: Finally, why didn’t the authors do a multivariate regression with all three components and report semi-partial correlation coefficients? I apologise I’m only asking this now in a 2nd round of review, but the comments from Reviewer 2 have prompted me to think more on this. They do show the R^2 of the simple regressions between each pair of variables and I don’t think overall the rank order of what is important will change. I also appreciate that PFT area is likely not correlated much with Aleaf or LAI (though those two components are likely are correlated), especially for models with prescribed PFTs so a simple linear regression suffices. But for models with dynamic vegetation PFT area could be correlated with Aleaf and LAI and so knowing the semi-partial correlation coefficients would be interesting. I think this also might help satisfy the readers who might be tempted to think the single variable linear regression is too simple? I think the multivariate regression would also take care of the problem of forcing the regression through zero in the PFT area vs GPP case.

Response: We thank the reviewer for the suggestion and have now included a multivariate regression analysis, where we used PFT Area, LAI, and Aleaf as predictors and reported semi-partial correlation coefficients (Table S5). As the reviewer and we expected, the rank order of variable importance is consistent with our main conclusion, that PFT Area is a main controlling factor of total GPP variation across models, followed by LAI and Aleaf. We have now included this new analysis in our manuscript (Table S5, L307-313).

“We found the variation in PFT GPP is more closely linked to variation in PFT area than that in LAI or leaf photosynthetic rates (Fig. S8a). Using other approaches (i.e., multivariate analysis and variance decomposition analysis), we reaffirmed that PFT area is more important than LAI and leaf photosynthetic rates in controlling the inter-model GPP variation for each PFT (Table S5). This suggests that biogeography is likely a more important factor influencing the global GPP than LAI and leaf biogeochemistry”.

To further consider the co-dependence of PFT Area, LAI and Aleaf in the regression analysis, we employed a relative importance analysis using an advanced variance decomposition approach (i.e., Shapley regression). This approach allows us to fairly attribute the total explained variance (R^2) to each predictor, and can effectively minimize the effect of variable scale or inter-correlations between variables. This additional analysis also confirms that PFT area explains the largest share of the inter-model GPP variations (Table S5). We are also happy to see that these additional results also point out that for certain PFT, such as C₃ grasses (C3Gw, C3Gk, and C3Gc), LAI seem to be a key factor of inter-model GPP variation.

Table S5. Contributions of PFT Area, LAI and leaf photosynthetic rate (Aleaf) to inter-model GPP spread quantified by multivariate regression analysis and variance decomposition analysis (Shapley regression).

PFT	<i>r</i>			Sharply values		
	Area	LAI	Aleaf	Area	LAI	Aleaf
EBF	0.839	0.372	0.361	0.672	0.077	0.056
ENFt	0.807	0.617	0.522	0.543	0.266	0.179
ENFb	0.868	0.416	0.246	0.678	0.096	0.032
DBF	0.890	0.230	0.189	0.872	0.047	0.054
DNF	0.856	0.015	-0.066	0.842	0.115	0.030
SHR	0.983	0.282	-0.013	0.936	0.043	-0.001
C3Gw	0.591	0.381	-0.223	0.272	0.162	0.066
C3Gk	0.633	0.494	0.021	0.483	0.338	0.035
C3Gc	0.641	0.348	-0.159	0.351	0.157	0.056
C3C	0.690	0.541	0.496	0.405	0.086	0.104
C4G	0.237	0.460	0.265	0.332	0.465	0.044
C4C	0.312	0.148	0.087	0.521	0.439	0.010
AVG	0.696	0.359	0.144	0.576	0.191	0.055

Note: *r* is the semi-partial correlation coefficients from the multivariate regression and Sharply values indicate contributions of each predictor to total explained variance. Large absolute *r* and Sharply values indicate a larger contribution from the factor. AVG is the average contribution from each factor to the spread of inter-model GPP across all PFTs.

R3C6: I also like Reviewer 2's suggestion about looking at the spread in GPP values across for locations where modeled PFTs correspond well to each other (e.g. low CV in Fig. S5) and the spread for locations where modeled PFTs do not correspond well to each other (higher CV in Fig. S5). I hope the authors take these concerns and suggestions in the spirit they are intended. As the authors note, very few people consider that PFT area is an important component of spread in total model GPP. But I fear that with the simple linear regression, readers may be quick to try to dismiss the claim that PFT area is indeed important.

Response: Thank you for the comment. We appreciate the concern from the reviewer that our main message might not be fully appreciated if our conclusion is based on a linear regression – we are perhaps less worried about that, considering the relationships were statistically robust and consistent with our hypothesis. We have followed the advice of the reviewer to generate a new figure demonstrating the spread of inter-model PFT GPP at the grid level (indicated by CV of GPP; Fig. R1). We further compared the CV of GPP to the CV of Area (Fig. S5), and found the two metrics show strong relationships (Fig. R2). This implies that at the pixel level, the inter-model variation in GPP simulation (CV of GPP) is also highly dependent on the inter-model variation in area simulation (CV of area).

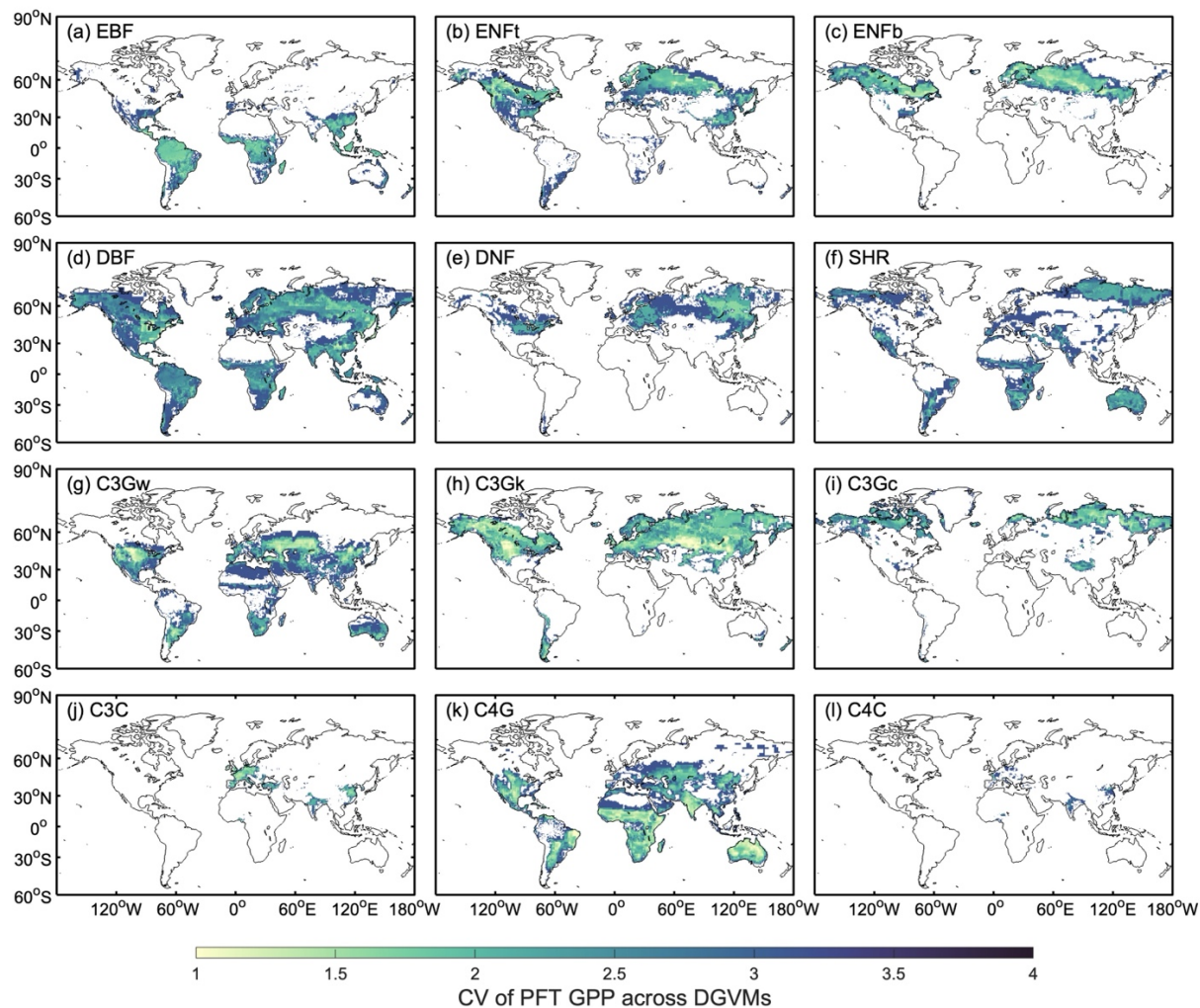


Figure R1. The difference in the distribution of PFT GPP in DGVMs. The coefficient of variation (CV) is calculated as the ratio of the standard deviation to the mean PFT GPP across DGVMs, with higher CV values indicating greater uncertainties in PFT GPP simulations from the ensemble of DGVMs.

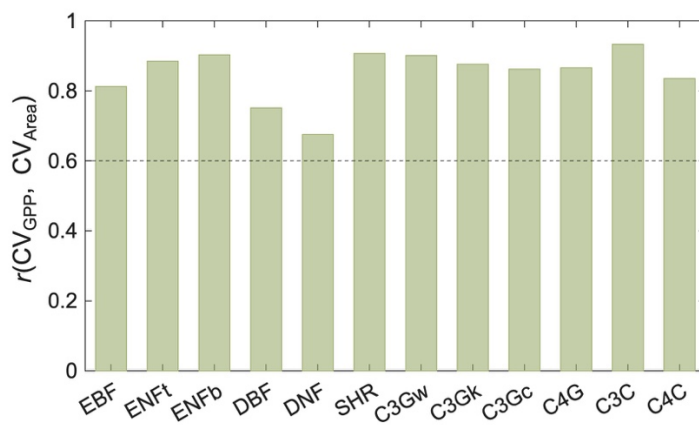


Figure R2. The correlation between CV of GPP and CV of Area across all pixels. The correlation coefficients are above 0.6 for all PFTs.

R3C7: I am more persuaded by the authors' argument related to the emulators that if you "correct" the PFT location (not just the area) and then get a lower reduction in uncertainty then you can postulate that is because the PFT parameters (or sensitivity to climate of that location) have been tuned to the wrong location. However, I'm not sure I'd phrase it as PFT locations were "correct" just because the authors have used the same prescribed PFT for all model emulators. Perhaps "consistent" would be a better word? Also, MODIS is not necessarily correct, and therefore couldn't it be the case that MODIS is incorrect, so now the models do not have the correct PFT in that location? Thus having a more incorrect PFT location by using MODIS could be contributing to the lack of reduction in uncertainty, unless I'm misunderstanding something? I like Fig. S6 (although note that it's not cited in the text), and I appreciate the extra analysis and response to this point, but there are not many models that have PFTs in regions we're not expecting. Looking at Fig. S5 the highest CV across models tends to occur at the boundaries of the typical ranges of each PFT, suggesting to me that it's more about slight shifts in distributions than completely wrong locations/different climate? What I will say is that this is nonetheless an interesting discussion point. If what the authors suppose here is true, then it has implications for how we optimise parameters – it suggests we should not be using RS maps to set PFT fractional cover in parameter optimizations. Even if RS data are used, the PFT fractional cover for that grid cells should be independently validated with field data. This in turn has implications for using complete gridded satellite data (for example all grid cells worldwide) to optimize parameters. This can be easily tested with two optimisations using differing PFT fractional cover and I'm not sure this has really been done before (not suggesting the authors do this at all – merely thinking on paper).

Response: We agree with the reviewer that 'consistent' is a more appropriate word than "correct" for the statement in L280 and have updated the statement to reflect that.

We fully concur that for models that use RS PFTs to prescribe their PFT maps, they might suffer from the uncertainty of RS PFTs in the simulations. We have added a statement in L251-254 to make this point more explicit.

"Additionally, when a remote sensing PFT map (e.g., ESA CCI) is used to prescribe PFTs in a model (e.g., ORCHIDEE), the uncertainty in the remote sensing map will inevitably propagate into the uncertainty of model simulations."

Regarding the uncertainty in DGVM PFT maps (Fig. S5), as the reviewers noted, the places with largest uncertainty in PFT may appear either on boundaries of PFTs or more broadly across the globe (e.g., DBF, SHR). We are currently working on another study trying to figure out the boundaries of various PFTs using a combination of remote sensing and biogeographical models, and it might be a bit premature to comment on the boundary issues at this stage (e.g.,

our preliminary results show that the competition not on boundaries can also impact PFT composition to a certain degree, while boundary is just a rough control on distribution).

From the modelling perspective on optimizing DGVM parameterization, we agree with the reviewer that using different RS PFTs likely give very different sets of optimized parameters for DGVMs, so for that line of study the modelers should be aware of the issue.

Another point which might be worth highlighting is the different magnitudes of uncertainties in RS PFTs (Fig. 3) and DGVM PFTs (Fig. S5). The disagreement in RS PFTs is still relatively smaller compared to DGVM PFTs. Therefore, it is still feasible to use RS to constrain land carbon modelling. We have added that statement in the discussion (L359-361) to explain it.

“Nevertheless, the uncertainty in RS PFTs (Fig. 3) is still smaller than models PFTs (Fig. S5), making it feasible to use RS PFTs to constrain the land carbon cycle in DGVMs.”

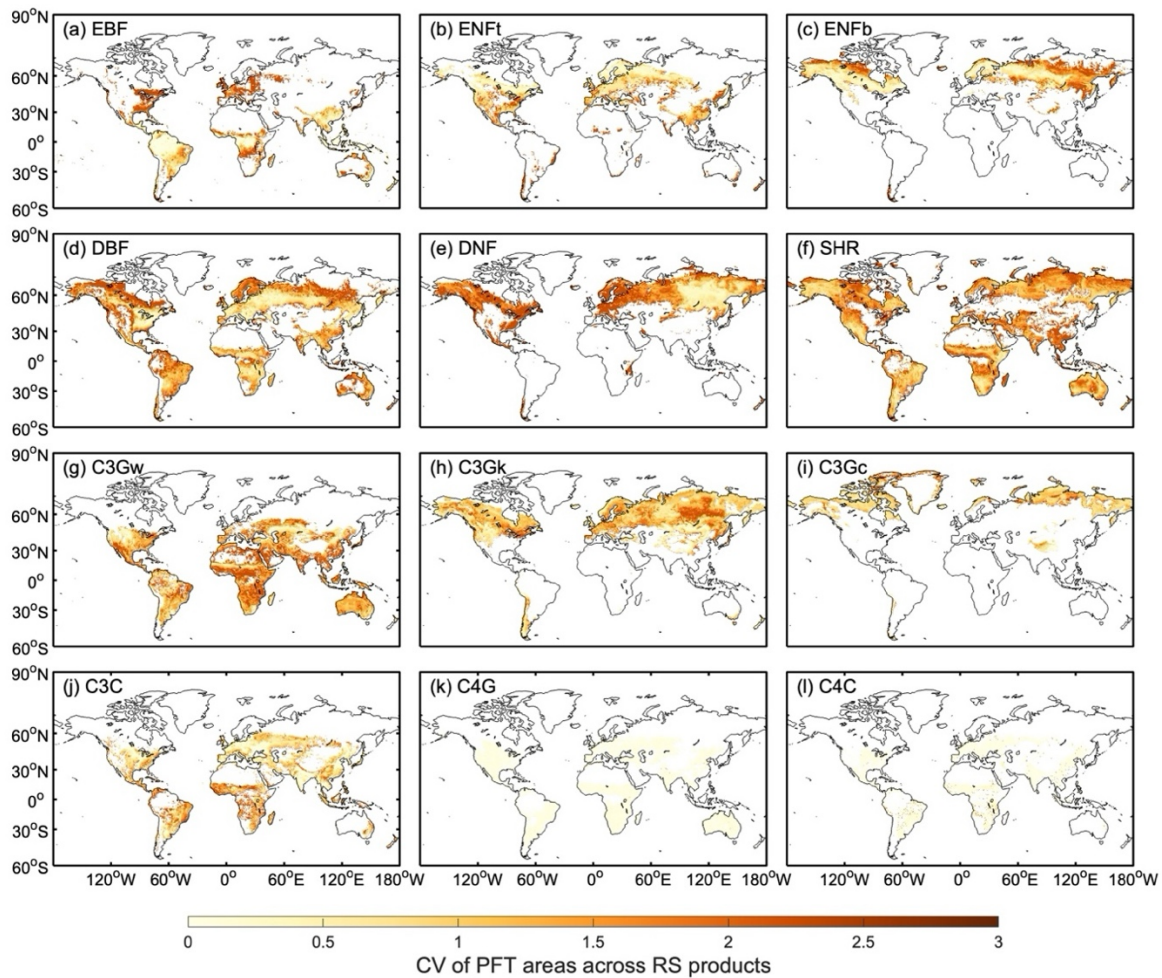


Figure 3. The difference in the distribution of PFTs from remote sensing-based (RS) products.

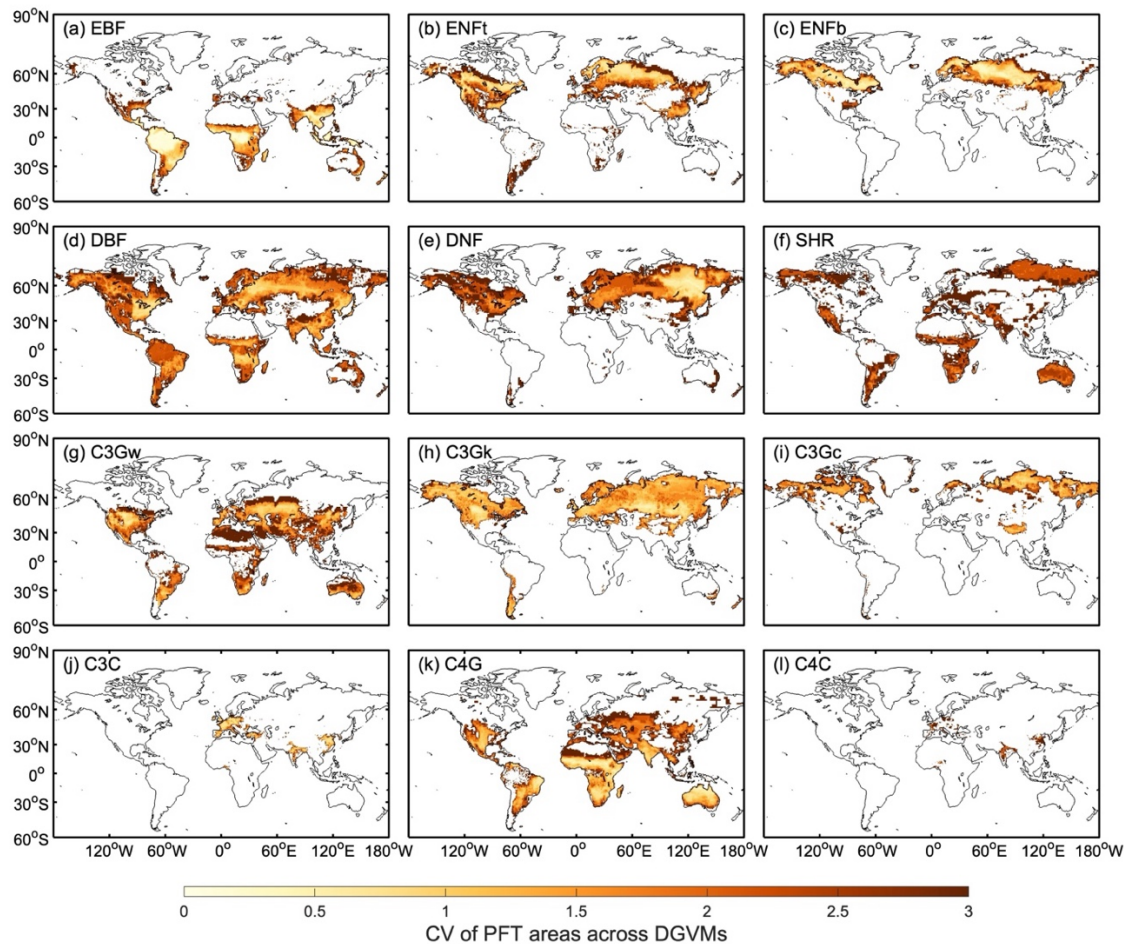


Figure S5. The difference in the distribution of PFTs in DGVMs.

R3C8: One more question about the emulators, just for clarification. Were the emulators were run for each grid cell and the pft fraction for that grid cell was mapped to the pft dependent GPP for that grid cell? Or was it scaled up so it's total GPP corresponding to total PFT area?

Response: Thank you for the comment. To further clarify, we ran emulators for each PFT, with the input of the emulator being climate for each grid cell and output being *GPP rate* ($\text{gC/m}^2/\text{year}$) for that PFT. After which, we used the emulators to simulate GPP for that PFT at the grid cell level. By multiplying GPP rate ($\text{gC/m}^2/\text{year}$) and PFT area (m^2) in that grid cell, and then summing them up across all grid cells, we obtained PFT-specific total GPP. The process is similar to DGVM structures. We added the text below to clarify this point (L521-524),

“After developing these emulators, we used them to estimate PFT-specific GPP rate within each grid cell and obtain PFT-specific GPP by using a consistent PFT fraction map from MODIS. We then aggregated the PFT-specific GPP of all grid cells and calculated global GPP”.

R3C9: There really seems to be an issue with crop cover in the models, if I'm not

misunderstanding? And a wide range across RS products for C3 crops. I'm surprised by this. Again, more thinking on paper.

Response: Just to support the reviewer's thinking, yes, there is a wide range in C₃ crop area among RS products, from 7.66 million km² (MODIS) to 16.34 million km² (ESA CCI). The disagreement is largest in the tropics (~ 30°S - 30°N) and parts of Europe (Fig. R3).

This discrepancy, from our preliminary analysis, might stem from different definitions and thresholds used for classification in each RS. For MODIS, cropland is defined as land where at least 60% of the area is cultivated. ESA CCI uses a broader definition for cropland, including Mosaic cropland (> 50%)/Natural vegetation (< 50%), rainfed and irrigated croplands. However, to what degree the difference in cropland definition contributes to the difference in area might need further examination.

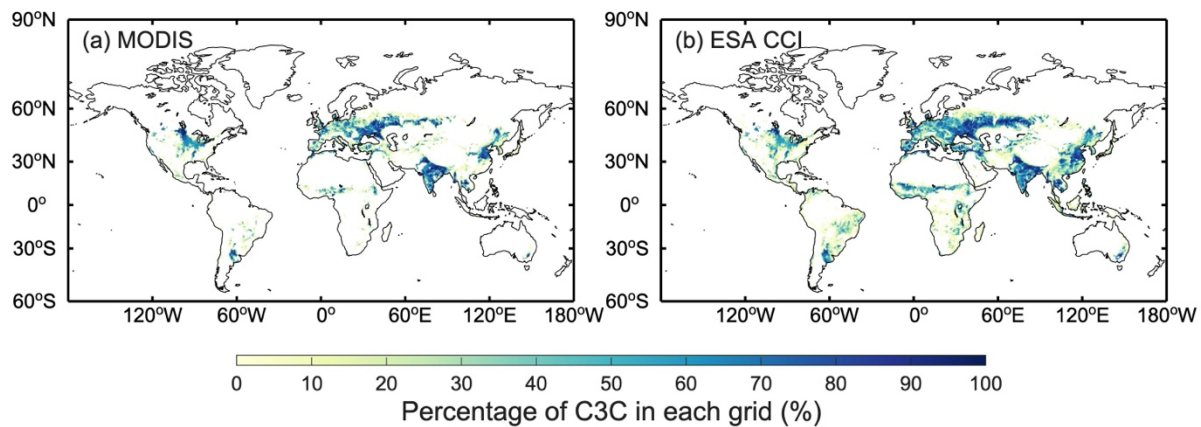


Figure R3. The spatial disagreement in C₃ crop (C₃C) area between MODIS (a) and ESA CCI (b).

R3C10: Line 204: did you mean to cite Fig. 3 here?

Response: Thank you for pointing this out and we have now cited the correct figures (Fig. S5 and Fig. S6) here.

Reference:

Rogers, A. 2014. The use and misuse of $V_{c, max}$ in Earth System Models. *Photosynthesis research*, 119(1), 15-29.

Anav, A., Friedlingstein, P., Kidston, M., et al., 2013. Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth System Models. *J. Clim.* 26, 6801e6843.

Song, X., Wang, D. Y., Li, F., et al., 2021. Evaluating the performance of CMIP6 Earth system models in simulating global vegetation structure and distribution. *Advances in Climate Change Research*, 12(4), 584-595.

Piao, S., Wang, X., Park, T., et al., 2020. Characteristics, drivers and feedbacks of global greening. *Nature Reviews Earth & Environment*, 1(1), 14-27.

Walker, A. P., De Kauwe, M. G., Bastos, A., et al., 2021. Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO_2 . *New phytologist*, 229(5), 2413-2445.