Responses to Anonymous Referee #1 on submission to Geoscientific Model Development Discussion (doi:10.5194/gmd-2016-22).

Submitted by Anna Harper on behalf of myself and my co-authors.

We thank the referee for your helpful comments on the manuscript and for taking the time to review it. Below we include the referee comments in black and our responses in red. The supplement contains a revised manuscript with red indicating changed sections. All line numbers refer to that version of the manuscript. Note the revised manuscript also includes some edits of minor errors (all in red for traceability).

The paper "Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information" present several improvements of the JULES DGVM. these improvement are first based on increasing the number of PFTs from 5 to 9 to better represent the different types of leaves in the leaf economic spectrum including deciduous and evergreen trees and a separation between climate zones. Second improvement was done in estimation of leaf photosynthesis from leaf nitrogen and improvement of phenology considering a more realistic leaf longevity.

This is an important paper that allow to follow recent developments of the JULES model and perfectly fit to the objective GMD. The changes are sufficiently important to justify the publication of a paper. The paper is well written with a convincing evaluation of new model performances both at site level and at global scale. The results show a clear improvement of the model at different scales. For all these reasons I recommend the paper for publication. Here after are just some minor comments that could help to improve the manuscript:

- There is no real justification of the choice of 9 PFTs except as a minimum to represent the main leaves forms. Obviously, for technical reasons, the number of PFTs cannot be increase indefinitely and then a compromise should be find but it would be interesting to see if including a higher number of PFT should also give higher performances?
One way could be to look to the differences between simulated GPP and NPP and respectively Jung and MODIS maps for each pixels and each PFT. Then we could see if there is spatially coherent systematic bias that could show possible new PFT separation.

It's true that the choice of PFTs is subjective. The 9 PFTs were chosen as they represent the range of deciduous and evergreen plant types with minimal externally determined bioclimatic limits. The distinction between tropical and temperate broadleaf evergreen trees exists to account for the important differences between these types of trees, as described in the Introduction. In particular, measured V_{cmax} for a given leaf N per unit area (N_A) can be lower in tropical evergreen trees than in temperate broadleaf evergreen trees (Kattge et al., 2009), resulting in lower V_{cmax} and maximum assimilation rates for tropical forests. We have added these justifications in the Discussion (Lines 759-763).

Previously JULES was hard-wired for 5 PFTs. An important step in going from 5 to 9 PFTs was removing this hardwiring. Now users can define the number of PFTs, so the 9 documented in this paper are a recommendation but can be adjusted in the future. This is now mentioned at Line 754-756.

However, it is a good suggestion to evaluate in a more objective way if an appropriate number of PFTs has been chosen. One logical way to further subdivide the PFTs is based on the biome maps, which is similar to the reviewer's suggestion. The analysis in the manuscript was based on an original data set of 14 biomes, where some biomes were combined for a total of 7 biomes. In a new figure (Fig. SM6) we show the biases in JULES5 and JULES9 for 11 of the original 14 biomes (3 biomes are very small and had no visible differences in the maps: Tropical Coniferous forests, Mangroves, and Flooded grasslands/savannas). The area-weighted RMSE is given in the top left of each map. Some biomes do not show an improvement in JULES9 and this gives some indication where extra PFTs might improve the simulation: for example the Boreal Forests/Tiaga; Tundra; Mediterranean woodlands; and Desert/Xeric Shrublands. Also the biases are still very high for the tropical/subtropical forest and grassland biomes. These regions broadly agree with what was mentioned in Section 5.2. We have added a more specific recommendation for development in these regions at Line 770-772.

However we also caution against defining too many PFTs, as there is already overlap between the N_m and LMA traits (Fig. 1c). In developing new PFTs it would be ideal to determine definitions that result in distinctive sets of traits. Future work will address the possibility of more PFTs or improved processes with the new framework for flexible PFTs in JULES.

minor comments on figures:
- Figure 7: what represent the grey zone ?
This is the standard deviation of the observed fluxes, based on the monthly means from all months. This information has been added to the caption.
- Figure 8: The figure is difficult to read mainly because this is tiny figures. Should it be possible to split it to have larger figures ?
We rearranged this figure so the individual panels can be larger.

Responses to Anonymous Referee #2 on submission to Geoscientific Model Development Discussion (doi:10.5194/gmd-2016-22).

Submitted by Anna Harper on behalf of myself and my co-authors.

We thank the referee for your helpful comments on the manuscript and for taking the time to review it. Below we include the referee comments in black (with the specific questions addressed in bold), and our responses in red. The supplement contains a revised manuscript with red indicating changed sections. All line numbers refer to that version of the manuscript. Note the revised manuscript also includes some edits of minor errors (all in red for traceability).

The authors present a version of the JULES land surface model with a more detailed dynamic vegetation model and show that this gives more accurate carbon fluxes than the traditional version of JULES. It is of great interest and should be published. **My only question is whether you could you have got the same answer by tuning the old version of JULES?** Adding extra PFTs will cause greater complication than tuning parameters, especially when competition between PFTs is turned on. You say you corrected known biases in the model. **Did these same biases get corrected in the original, 5 PFT version, or just the new version?** If not, I think you should have added an extra experiment to assess the relative impact on the flux from adding the additional PFTs and the tuning. **Would just correcting the 5 PFT JULES have had the same impact as adding extra PFTs?** I think that some discussion of this, and ideally an extra experiment, is needed.

First, we address the question of tuning. In this study, we have used observations to constrain the model. This has improved the model and it has helped detect areas of the model that are wrong and require further improvements to representation of processes. The parameter changes that have been made are backed up with data and so we are putting the right values for the right reason. Tuning can give you the right answer but not always for the right reason, and so should be done carefully.

There is ongoing work to tune certain JULES parameters. Another paper is in review with GMD to evaluate the tuning method (Raoult et al., *in review*). The next step in the model's development will be to combine the tuning with the new trait-based representation presented in this study.

We argue that the extra complication that results from the new PFTs is worth the benefit of having more diverse plant types, which should enable more diverse and specific responses to climate change. A follow-up paper is being finalized which analyzes the impacts of the new PFTs when JULES is run with dynamic vegetation, and results are also improved in this mode.

At the same time, it would be good to evaluate the improvements with extra PFTs compared to just improving parameters with 5 PFTs. As the reviewer suggested, we added a third global experiment to test the 5 PFTs with improved parameters, as in Table SM2. The supplemental material now includes this table plus recommendations for running JULES with 5 PFTs and improved parameters.

Table SM2. New trait-based parameters for 5 PFTs that are consistent with TRY data. N_m, LMA, and γ_0 (=1/[leaf lifespan in years]) were calculated directly from the data collected. The slopes and intercept parameters for V_{cmax} (s_v and i_v, respectively) were calculated based on the average of observed values available from Kattge et al. (2009).

	BT	NT	C3	C4	SH
N_m	0.0185	0.0117	0.0240	0.0113	0.0175
LMA	0.1012	0.2240	0.0495	0.1370	0.1023
s_v	25.48	18.15	40.96	20.48	23.15
i_v	6.12	6.32	6.42	0.00	14.71
$V_{cmax,25}$	53.84	53.88	55.08	31.71	56.15
T_{off}	5	-40	5	5	-40
d_T	9	9	0	0	9
γ_0	0.25	0.25	3.0	3.0	0.66
γ_p	20	15	20	20	15
L_{min}	1	1	1	1	1
L_{max}	9	7	3	3	4

We also change the following parameters from their default value in Table 1 to make the parameters consistent with $JULES9_{ALL}$:

	BT	NT	C_3	C_4	SH
D_{crit}	0.09	0.06	0.051	0.075	0.037
f_0	0.875	0.875	0.931	0.800	0.950
f_d	0.010	0.015	0.019	0.019	0.015
rootd	3	2	0.5	0.5	1
T_{low}	5	0	10	13	0
T_{opt}	39	32	28	41	32
T_{upp}	43	36	32	45	36
α	0.08	0.08	0.06	0.04	0.08
μ_{rl}	0.67	0.67	0.72	0.72	0.67
μ_{sl}	0.10	0.10	1.00	1.00	0.10

The new experiment is called $JULES5_{ALL}$, since it included as many parameter updates as possible to give a fair comparison between JULES with 5 PFTs and $JULES9_{ALL}$. Most of the differences in GPP and NPP between $JULES5_{ALL}$ and $JULES9_{ALL}$ were in the tropics. The global GPP was high (135 Pg C yr^{-1}) in $JULES5_{ALL}$, primarily because V_{cmax} for the average broadleaf tree (53.84 μmol m^{-2} s^{-1}) was much higher than for the tropical broadleaf evergreen PFT (41.17 μmol m^{-2} s^{-1}). Although tropical GPP was higher in $JULES5_{ALL}$ compared to $JULES9_{ALL}$, the NPP was lower and closer to the values from MODIS NPP. The reason was the differences in leaf nitrogen, which increased respiratory costs in $JULES5_{ALL}$ compared to $JULES9_{ALL}$. Both N_A and N_m were higher for the broadleaf tree PFT (1.87 g N m^{-2} and 0.0185 g N g^{-1}, respectively) than for the tropical evergreen broadleaf tree PFT (1.77 g N m^{-2} and 0.0170 g N g^{-1}, respectively).

We have added an explanation of this simulation, its results, and implications in the manuscript at Lines: 439-442, 679-687, and 759-767. Also the global results are shown in a new figure (Fig. 9) and summarized on a per-biome basis in Table 6.

Table 6a. Area-weighted GPP from each biome (g C m^{-2} yr^{-1}). The biome total GPP from MTE is given in Pg C yr^{-1} to give perspective of each biome's role in the global total.

Biome	JULES5	JULES9	JULES5-ALL	MTE	MTE total
Tropical forest	2403±217	2295±191	2505±217	2244±297	49.9
Tropical forest: Only BET-Tr.	2924±144	2955±147	3279±178	2790±273	
Tropical savannah	1355±244	1268±223	1320±237	1111±257	21.9
Extratropical mixed forests	947±147	1082±158	1119±167	1119±212	2.9 (13.4*)
Boreal and coniferous forests	514±99	597±118	645±122	650±203	12.1
Temperate grasslands	420±145	465±138	477±140	509±184	8.1
Deserts and shrublands	82±48	91±46	91±47	283±200	4.9
Tundra	86±20	94±20	101±20	279±233	1.9
Mediterranean Woodlands	324±147	407±136	405±140	510±190	1.5

*Value for EMF biome when agricultural mask is not applied.

Table 6b. Area-weighted NPP from each biome (g C m^{-2} yr^{-1}).

Biome	JULES5	JULES9	JULES5-ALL	MODIS17
Tropical forest	956±144	1007±125	951±143	786±352
Only BET-Tr.	1141±101	1233±103	1109±126	929±315
Tropical savannah	527±158	591±143	584±152	451±319
Extratropical mixed forests	586±93	631±104	640±110	563±231
Boreal and coniferous forests	307±65	358±77	385±80	350±155
Temperate grasslands	180±94	243±89	242±90	304±247
Deserts and shrublands	16±29	35±29	33±29	111±133
Tundra	52±14	61±13	65±13	136±94
Mediterranean Woodlands	118±94	201±89	195±89	324±184

Further referee comments:

Experiments 4+ are discussed before experiments 1 to 3 in the text. It would be easier to follow if all the experiments were described in the same way and in the same order. Perhaps move the method around line 515 from the results section to before the first mention of experiment 4?

We switched sections 2.3.1 and 2.3.2 so the Experiments are described in Section 2 in the correct order. We also added further explanation of these experiments at the beginning of Section 2 (~Line 170-172), and of the calculation of the relative statistics (Line 469-470). However now Table 4 is mentioned before Table 3 so these are switched throughout the manuscript.

Table SM 2 gives tuned parameters for the tuned 5 PFT JULES, but I cannot find a reference to that in the text. Is there a missing section?

Yes Table SM2 should be referenced in the text. Thank you for catching this. It is now referred to at Line 176-177. Also extra discussion is added to the supplementary material (see page 3 of SM).

"and updated the model phenology to include a trade-off between leaf lifespan and leaf mass per unit area." - Does your improvement not just change the leaf turnover rate and its impact on the carbon flux rather than the phenology, which is still controlled in the same way as traditional JULES?

It is true that the equations controlling phenology in JULES (Eq. 15-16) were not changed. However, changing the temperature threshold, T_{off}, did change the timing of when leaves grow in the fall and senesce in the fall. The trade-off referred to here is included in JULES by increasing leaf growth in the spring (γ_p) and turnover rates in the fall (γ_0) for leaves with low LMA, while maintaining low turnover rates for the thicker, longer-lived leaves. However it could be misleading to say the phenology was updated since no structural changes were made to the model so we have reworded this sentence in the abstract.

General comment for reviewers:

Note that in two places we have changed “tuning” to “calibration” as the parameter changes were not really tuned in a strict sense (Line 167, Line 1204). There is a tool for tuning parameters in JULES (adJULES, Raoult et al., 2016), but this was not used in this study. So we believe the change from “tuning” to “calibration” is a more appropriate description of what was done, and will avoid confusion between what can be done with adJULES and the techniques used in this study (adjustment of parameters to correct biases, or more frequently new parameters based on data and literature review). The justification for each parameter change has already been provided in the Methods section.

Lines 167-168: Updated parameters were based on review of literature