

Interactive comment on “The importance of tree demography and root water uptake for modelling the carbon and water cycles of Amazonia” by Emilie Joetzjer et al.

Anonymous Referee #1

Received and published: 26 September 2018

The manuscript by Joetzjer et al. explores how two new formulas for the ORCHIDEE dynamic global vegetation model influences carbon and water budget predictions of three Amazon forests. ORCHIDEE-DGVM, called TRUNK in this manuscript, is used as the baseline model and represents the land surface physiology as an aggregated ‘big leaf’. The first new formula, ORCHIDEE-CAN (CAN), represents tree demography and aboveground mechanistic water transport through the stem, and is parameterized specifically for Amazon forests. The second new formula (CAN-RS) modifies CAN to include mechanistic soil water uptake by the roots. Improving the representation of belowground processes in land surface models is at the forefront of discussions in the modelling community and is a recommendation in the recent paper by Fisher et al.

C1

(2018), which is an authoritative review paper about the state-of-the-science of second generation dynamic vegetation models. I read this manuscript with great enthusiasm since these two formulas, CAN and CAN-RS, are at the cutting edge of model development needs. I do feel that this manuscript has great potential to add to our scientific knowledge about individual tree water-use strategies and physiology. However, in my opinion this manuscript is still too premature to be considered for publication. I support my opinion with the following comments.

General comments.

1. My main critique of this manuscript, and one that I feel should be sufficiently addressed before it is considered for publication, is that it presents three alternative hypotheses about biological controls over tropical forest carbon and water cycles, yet these hypotheses are not clearly set up in the Introduction, described in the Methods, or convincingly interpreted in the Results and Discussion (comments on each section are detailed below). Also, given the results, the manuscript does not convincingly justify the asserted need to increase the TRUNK model complexity specifically with the new CAN-RS formulation.

Introduction.

(a) I recommend eliminating or reducing the first paragraph to a single sentence in order to create more space to describe the three hypotheses being evaluated.

(b) It is not clear what hypothesis TRUNK represents that is relevant to this study. I presume it has something to do with spatial aggregation; but since there is only a single line (P3, ln7) written about TRUNK, it is unclear what CAN and CAN-RS are being compared to. It would be helpful to have some context about where TRUNK has been a good hypothesis for explaining forest function, and carbon and water exchanges, and in what cases the hypothesis is not well supported.

(c) I feel that the Introduction will flow better if the three hypotheses that are all jointly

C2

introduced at the end of the Introduction are instead introduced in each's paragraph that contains its respective problem statement. For example, in paragraph three of the Introduction, (P2, ln23-29), it would be clearer if the proposed remedy (i.e. increased "layer-to-layer heterogeneity in soil-to-root resistance) is included as the concluding sentence of this problem-statement paragraph.

(d) It would be helpful to articulate in the Introduction what "fingerprints" the data are expected to have that would support of each particular hypothesis. E.g. what fingerprint (or marker) should be in the SWC data to support the given dynamic-root-water-uptake hypothesis? (Conversely, if dynamic root-water uptake is not important, what should the SWC data look like?) Articulating the fingerprints gives a justification for the selection of each benchmark presented. As the paper is currently written, the reader is first introduced to the specific benchmarks in the Results and has to intuit the rationale for each.

Methods.

(a) Eqns. 3,4, 5 and 6 need to include an index (e.g. i) for each variable that is calculated separately for each soil layer (for example, if $\psi_{s,i}$ at P5 ln10 is calculated by layer, it should be $\psi_{s,i}$, and i needs to be defined). Similarly, it is not clear how the model differentiates between the resource-use of different-sized trees. Should there be an index (e.g. j) in Eqn 5 for each size class of trees? Or, are all the roots uniformly distributed such that the seedlings have the same rooting depth and access to water as the canopy trees? Please clarify how this works. If it is the latter, then the new hydrodynamic physiology is not mechanistically linked to demography.

(b) I feel that the language related to the demography hypothesis needs to be much more precise. It is important to distinguish between whether the demographic rates emerge or are prescribed. If the demographic rates emerge due to a mechanism, then what is the hypothesized mechanism (C-starvation, hydraulic failure, etc.)? If the demographic rates are prescribed by the user, then this model does not simulate self-

C3

thinning through competition for limiting resources (P4 ln18) and it does not simulate demography as suggested in the Methods. Rather, the model represents the outcome of competition and self-thinning by prescribing a number-density by size-class distribution. In other words, does the model neglect all of the size-dependent internal dynamics of the ecosystem that give rise to the demographic rates from the bottom-up? If that is the hypothesis, then it needs to be clarified. Eqn 5 will act differently on individuals of different sizes due to non-linearities, than it will on all of the roots aggregated together. Making this distinction has very important implications for how ecologists and physiologist interpret the results of the simulations.

(c) The soil-water stress function for TRUNK needs to also be included. Is it similar to Eqn 3?

(d) I recommend that all parameters that are new or have new values introduced in this study be placed in a Table that includes the appropriate reference. It is a little cumbersome to have to search the text for this information.

Results and Discussion

(a) I am not convinced by the interpretation of the results for CAN and CAN-RS in terms of the claims that this paper is trying to make. The title implies that demography (CAN) and root water uptake (CAN-RS) are important for modelling tropical forest carbon and water cycles. The Abstract claims that modeling root water uptake in greater detail (CAN-RS) improves model performance (P1, ln32-35). The Results, Discussion and Conclusions make similar claims throughout. I do not feel that these conclusions are supported by the evidence presented in the Results for two primary reasons.

First, the Results need to give a much more thorough description of the Taylor Diagram (Figure 3); three sentences is not adequate (P10, ln8-12). Contrary to the Title and Abstract, the Taylor Diagram in Figure 3 indicates that TRUNK is an equivalent, but likely better, predictor of LE, GPP and NEE than CAN or CAN-RS. This is a highly significant outcome, yet it seems to be downplayed. In all three panels, TRUNK has

C4

better or equivalent RMSE and correlation scores compared to the two other models. CAN and CAN-RS seem to only score better than TRUNK for the standard deviation metric. RMSE and correlation are indicators that the model is getting the pattern, and hence the mechanism, correct, whereas standard deviation indicates how well the mechanism is tuned to correctly capture the magnitude of the response. Therefore, for this type of analysis, which explores mechanistic controls, RMSE and correlation are more important indicators of model performance; yet the Results focus primarily on standard deviation. One objective way forward is to assign a skill score to each model as proposed by Taylor (2001) in Eqn 4 or 5. If a “penalty” is imposed, as in Eqn 5 of Taylor (2001), then the justification for the penalty needs to be described in the Methods.

Second, in just about all cases, TRUNK appears to be a better predictor of the benchmarks than either CAN or CAN-RS (Figures 3, 4, 5). To begin with, TRUNK seems to match the observations in Figure 4 the best of the three models. CAN seems to be the weakest predictor, which tells me that either demography itself is not important, or the hypothesis contained in CAN about how to represent demography is not supported. Which is it and why? Next, both CAN and CAN-RS show a midday suppression in GPP during the dry season that is not present in the observations (Fig. 5); however, TRUNK does not show the midday suppression. This tells me that the mechanistic water transport hypothesis as it is represented in CAN and CAN-RS is not supported by the observations. In Figure 6, neither CAN nor CAN-RS reproduce observed SWC in any credible way. I am curious about what fingerprint the authors would expect the data to have in support of the CAN-RS hypothesis (this should be mentioned in the Introduction). Does Fig.6b actually possess such a fingerprint? On the other hand, the observations presented in Fig. 5 do not possess the expected fingerprint (i.e. midday suppression). If the observations do not possess the expected fingerprints, then what is the rationale for including the CAN-RS hypothesis in TRUNK in the first place? And, what is the rationale for using the eddy flux and SWC data as benchmarks if they do not contain the relevant fingerprints? I do wonder if the observed SWC data presented

C5

in Fig. 6 has the correct resolution to be a valid test for the CAN-RS hypothesis. If not, then the authors might consider Miller et al. (2011), it contains the fingerprint that supports the CAN-RS hypothesis (Fig. 1b) and perhaps would be a better test of the CAN-RS hypothesis. Finally, given that TRUNK has the correct pattern (and no midday suppression) but the incorrect magnitude in Figure 5, could it be that that the structure of TRUNK is correct, but it just needs better tuning?

(b) The Taylor Diagram is certainly a good method for cross-model comparisons to observations, but one of its limitations that should be addressed in the Discussion is that, unlike AIC for statistical models, it does not account for the trade-off between model simplicity and complexity. As acknowledged in the Introduction (P3, In3), process-based land surface models have extremely high degrees of freedom that can lead to compensating errors and equifinality. It is unclear if and how this trade-off has been considered in the conclusion that CAN-RS is an improvement over TRUNK.

2. This manuscript would be much more impactful with greater synthesis with the many recent and important advances in the field of dynamic vegetation model development. The manuscript also needs greater synthesis about how these different hypotheses inform our physiological and ecological understanding of tropical forests.

(a) For example, Xu et al. (2016) and Powell et al (2018) also use a dynamic vegetation model that explicitly represents demography and mechanistic water movement through the soil-plant-atmosphere continuum. Xu. et al (2016) explores mechanisms related to water-stress avoidance, while Powell et al. (2018) explores mechanisms related to tolerance of water-stress. There is a good opportunity to connect the insights into belowground mechanisms explored in this study to the insights into aboveground mechanisms explored in those two studies.

(b) This study also addresses some of the issues raised in the recently published Anderregg et al. (2018).

(c) Figure 15. Figures of the spatial distribution of AGB across the Amazon basin have

C6

been widely published. Zhang et al. (2015) reported such figures for several DGVMs (Fig. 3s) including one like CAN that contains a size-structure and demography hypothesis. I feel that Figure 15 is a significant result because the CAN hypothesis does not show a strong spatial gradient in AGB, which is in contrast to two hypotheses in Zhang et al. (2015) that do capture the gradient reasonably well (see Figure S3f in Zhang et al.). Why is the CAN hypothesis not supported by the data, but the other two hypotheses are supported by the data? Is it due to CAN model structure, experimental design, parameterizations? What can we learn about the ecology and model development needs from the contrasting results?

(d) Finally, this study and Levine et al (2016) proposed two different hypotheses about how to represent demographic processes in a land surface models; and, the two studies produces very important and contrasting results. Levine et al. (2016) argues that the size-structured model hypothesis is supported by the data because the demographic rates emerge from a bottom-up formulation of spatial heterogeneity. The CAN formulation, in contrast, appears to be a top-down scaling hypothesis. Given that the bottom-up approach agrees with benchmark tests (e.g. Xu et al. 2016, Zhang et al. 2015, Powell et al. 2018), but the top-down approach (CAN) does not perform better than TRUNK (a spatially aggregated hypothesis), what does this tell us about these approaches for representing demographic processes?

Making these connections will provide a much more complete picture about the contribution this study makes to the state-of-the-science regarding demographic dynamic vegetation models, modeling plant hydrodynamics, and understanding tropical forest ecology in general.

3. The total number of Figures could be reduced. The manuscript loses focus between Figures 10 and 17 and the rationale for these figures is not very well established in the Introduction (i.e. I do not see anything regarding predictions of spatial variability in the Introduction). I recommend this paper focus on figures that directly test the specific demography and root water uptake hypotheses contained in TRUNK, CAN,

C7

and CAN-RS. Figures 10 to 17 are summaries and their relevance is predicated on the CAN and CAN-RS hypotheses being more strongly supported by the data than the TRUNK hypothesis. Also, the Discussion needs to be better integrated with the figures presented in the Results. All key results figures should be cited in the Discussion; if they are not, then this tells me that they are not central to the story of the manuscript and should be moved to the Supporting Material.

Specific comments.

P1. Line 1. Title. The title highlights the importance of tree demography, yet the experiment is not set up to explicitly isolate this, as CAN, contains at least three differences from TRUNK: (1) explicit tree demography, (2) mechanistic stem water transport, (3) Amazonian specific parameterization. With the CAN versus TRUNK comparison, the authors need to evaluate how each of these differences (as well as any others that may be unmentioned) individually impact model predictions. Otherwise, attribution to any particular modification is confounded by its interaction with the other two modifications.

P1. Line 22. "remains challenging". This is very vague and will mean many different things to many different people, much of which is unrelated to the subject of this manuscript. This needs to be clarified to keep the reader focused on the subject of this paper.

P1. Line 23. "These" lacks an antecedent and therefore it is unclear what the limitations are.

P1. Lines 25-28. Why evaluate these three hypotheses? The Abstract lacks a problem statement with sufficient detail to motivate their evaluation.

P1. Line 31. "as well as TRUNK...at local and regional scales..." If this is the case, then why use a more complex model? This statement also does not support the conclusion in line 34, "...improves the representation of biogeochemical cycles..."

P1. Line 35. Last sentence. This sentence is of course true...about a lot of things.

C8

However, this sentence is quite vague and therefore loses any specific meaning with respect to the key findings of this study. It should be revised to specify “the variation [in what aspect of] ecological functioning”. Plant hydraulics? Soil hydraulics? Life history strategies? Phenology? Biogeochemical cycling? Life forms—palms, shrubs, trees, lianas? Disturbance regimes? Etc.?

P2. Line 1. “will likely”. This is too strong of a statement. Perhaps substitute “ are predicted to”.

P2. Line 8. The “large-scale dieback” is a quite dated result that has been updated widely in the literature over the subsequent decade and a half. Consider revising to be more current with our understanding of the system.

P2. Line 20. Xu et al. (2016) should be referenced here.

P5. Eqn 4. Is there a typo? Change min to max?

P10. Table 2. AGB. I think it is worth noting that none of the models capture the observed spatial variability of AGB.

P15. Line 16. Exchange “slightly” with “by 33%” (the two are probably not equivalent).

Figure 9a. y-axis. Delta of what? Caption needs to define this delta.

Figure 9b. What is Nb? Caption needs to define this.

P20 Line 19. “. . .the model better captured the seasonality of GPP and LE.” This needs to clarify that this is a comparison between CAN and CA-RS, I presume. However, this statement does not necessarily appear to be true for CAN-RS versus TRUNK, which needs to be stated. But suppose that it is true that CAN-RS is a better predictor of seasonality than this parameterization of TRUNK, then could TRUNK be improved with better tuning? Does the TRUNK soil water stress function contain a parameter like m_{psi} in Eqn 3, and could this parameter be tuned to shift the strength of where the soil water stress occurs? If so, the implications of this hypothesis should be discussed

C9

relative to the CAN-RS hypothesis and our understanding of the physiology.

P21 Line 34. Fig 18a. Should it be Fig17a?

P22. Line8. “functional diversity needs to be accounted for.” This no longer needs to be a suggestion as Xu et al. (2016), Powell et al. (2018) and Anderegg et al. (2018) demonstrated this.

P22. Figure 17b. AGB versus precipitation has also been presented in Good et al. (2011) and Levine et al. (2016). The novelty of 17b is that the pattern is reproduced with a different hypothesis about how to represent demography and plant water stress compared to the other two studies. That is an important result and the differences between the hypotheses should be highlighted to inform our thinking about the ecology of this region.

Supplemental Table B1. What is PFT2?

Supplemental Table B1. K (recruitment parameter) is not in Eqn. 2 on page 4.

References:

- Anderegg et al. (2018) *Nature*. doi: 10.1038/s41586-018-0539-7
Fisher et al. (2018) *Global Change Biology*. doi: 10.1111/gcb.13910
Good et al. (2011) *Journal of Climate*. doi: 10.1175/2010JCLI3865.1
Levine et al. (2016) *PNAS*. 10.1073/pnas.1511344112.
Miller et al. (2011) *PNAS*. doi: 10.1073/pnas.1105068108
Powell et al. (2018) *New Phytologist*. doi: 10.1111/nph.15271
Xu et al. (2016) *New Phytologist*. doi: 10.1111/nph.14009
Zhang et al. (2015) *Global Change Biology*. doi: 10.1111/gcb.12903

C10

