

RC1

This is a very good paper that describes a new stomatal conductance model and compares it with two well-established models using eddy covariance data at two contrasting sites (coniferous and deciduous forest). The paper is well written, the model is adequately described, and the simulation protocols are appropriate. My comments are mostly intended to clarify the analyses and improve the discussion.

Dear Reviewer,

Thanks for your appreciation and valuable suggestions. We have revised our manuscript carefully based on the reviewers' comments. Please see our detailed responses for the changes we have made.

1. Figure 1. The drawing shows the leaf layers at the top of the stem (and this is explicitly stated on lines 93, 99-100), but the model is multilayer in its radiative transfer. This means that profiles of light can drive profiles of stomatal conductance. But if each leaf layer is at the same height in the canopy, does each layer experience the same gravitational potential and same hydraulic conductance (i.e., the path length for water flow is the same for all layers)? A statement to this is needed, and what is the implication of this assumption?

The short answer is yes. The gravity term is accounted for through the branch system, and Δz can be set to be zero or even negative at each branch (this is done during the initialization step). We have clarified in the text and figure caption that the height of each organ can be customized. See our changes in

- Lines 103-105 (revised text, clean version; hereafter) “By default, we accounted for gravity in root and stem (gravity not accounted for in leaves), and thus each canopy layer has its own gravitational pressure drop. Yet, the gravity correction can be customized by setting the height changes of each root and stem.”
- Figure 2 caption “We account for gravitational pressure drop in root and stem (not in leaves) in the example; however, gravitational pressure drop can be customized by setting the height change of each root and stem.”

2. Figure 5 should be discussed in more detail. The three stomatal models have quite different values for maximum stomatal conductance (at low VPD and high soil water potential). This is a fundamental difference among models, which then should affect the estimates of V_{cmax25} and K_{max} obtained from the inversion.

The model setups do impact stomatal conductance for the same input parameters. As BBM and MED models are used with a tuning factor on V_{cmax25} , stomatal conductance and thus photosynthetic rate ought to be lower than OSM. This was the reason for why maximum stomatal conductance differs among the three models and why fitted V_{cmax25} has to be higher in the empirical models. We clarify this in the main text along with Figure 5 in the revision:

- Lines 211-219: “As BBM and MED models were used with a tuning factor on leaf photosynthetic capacity (represented by maximal carboxylation rate and maximal electron transport rate at a reference temperature, V_{cmax25} , and J_{max25} at 25 °C, respectively), effective V_{cmax25} used to compute photosynthetic rate was lower in BBM and MED models compared to OSM (when the three models used the same inputs). As a result, BBM and MED model predicted stomatal conductance and photosynthetic rate should be lower than OSM (when the same model inputs were used; Fig. 6a,c). Further, if the models are fitted to the same dataset, BBM and MED tend to have higher fitted V_{cmax25} to compensate for the negative effect from the tuning factor. The three models

also differed in their sensitivity to soil moisture as the penalty for OSM increased with transpiration rate, whereas V_{cmax25} would not be downregulated at relatively wet soil (e.g., soil water potential > -1 MPa; Fig. 6b,d).”

3. Figures 7, 8. Why is NEE used for comparison with observations instead of GPP? I understand that GPP is a derived product whereas NEE is a direct measurement. However, NEE requires ecosystem respiration, which is obtained by fitting Eq. (9) to the NEE measurement. It is not surprising then that all 3 models do a good job at simulating NEE (in contrast to ET)?

Since we did not know how well the partitioning algorithms represented true GPP, we preferred a fitting of remaining respiration, which is an offset in NEE. We believe it is better to compare the model simulations to direct measurements. Also, fitted remaining respiration rates from the three models agree in their magnitudes for the tested three models, as it was fitted using the nighttime respiration.

4. Line 264-269. A more thorough discussion of the fitted values for V_{cmax25} is needed. The fitted values vary with stomatal model. BBM and MED have similar values, but OSM has a much lower value. The explanation (that this results from the $\beta_w = K/K_{max}$ tuning factor) is inadequate. Also, how realistic are the fitted values? The values cannot be compared directly with leaf estimates, but the values (which represent a bulk canopy of leaves) are actually comparable to (or even smaller; OSM) that representative leaf values of 40-60 $\mu\text{mol}/\text{m}^2/\text{s}$ found in leaf trait databases. This suggests almost a one-to-one scaling from the leaf to the canopy. Another point to discuss is that the estimated V_{cmax25} is only appropriate for a specific stomatal model, meaning that the land surface model always needs to be recalibrated if the stomatal model is changed. The authors never acknowledge this point.

Following reviewer’s comment 2, we have added a more detailed description to the theoretical model comparison, and highlighted how canopy level water loss and carbon gain may differ among different model setups when the same suite of input parameters were used. Since we did not have the V_{cmax} for the tested sites, we were not able to compare the fitted V_{cmax} to observations (If we do have the site level observations of V_{cmax} , we would use them as model input directly rather than fit these traits). However, according to Timothy Tomaszewski & Herman Sievering (2007), V_{cmax} for spruce ranged from 8-12 and J_{max} ranged from 46-57. And OSM fitted V_{cmax25} was indeed equally low (around 15). The V_{cmax25} fitted for BBM and MED were not representative because of the tuning factor. The effective V_{cmax25} used to calculate photosynthetic rate needs to be lower. As to the last point, we highlighted it in our revision

- Lines 382-385: “Given the under-performance of empirical models when we used a different tuning factor algorithm (on photosynthetic capacity), we highlight it here that (1) inverted model parameters to use in LSMs vary with the model used to fit these parameters, and (2) using parameters inverted from one model setup in another model would likely result in biases in model outputs.”
- Lines 304-309 “Given that the fitted parameters were bulk properties of the sites, we expected them to differ from leaf-level observations but be of the same magnitude. However, because of the limited direct measurements in the studies forest sites, we were only able to find one study reporting a V_{cmax25} ranged from 8 to 12 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and a J_{max25} ranged from 46 to 57 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at Niwot Ridge (Tomaszewski and Sievering, 2007). Therefore, the OSM estimated $V_{cmax25} = 15 \mu\text{mol m}^{-2} \text{s}^{-1}$ seemed to be reasonable; and as we explained, BBM and MED estimated V_{cmax25} had to be higher than the OSM estimate due to the tuning factor.”

5. Line 281-290. All 3 models have similar K_{max} at the coniferous site, but quite different values at the deciduous site. The discussion relates this to differences in the C parameter in the Weibull function. How realistic are the fitted values for B and C in the function for the deciduous site?

The Weibull B and C parameters for the Ozark were from observations (not fitted by our model) as reported in Table 2. There is a long lasting debate about whether an exponential xylem vulnerability curve makes sense for plants. Yet, as seen from the OSM model performance and BBM and MED model performances (when we fitted g_1), the Weibull B and C used in this study seemed to be reasonable; otherwise, none of the models would perform well enough.

6. Line 322-331. A more thorough discussion of the prescribed (CLM) values for g_1 in BBM and MED and the fitted values is needed. The fitted values in Table 3 (BBM at Niwot and MOFLUX; MED at MOFLUX) are much higher than in CLM and much higher than estimates based on leaf gas exchange. Also, the fitted values for V_{cmax25} are lower (and more comparable to the values for OSM) when g_1 is fitted. This, again, tells me that fitting a model to eddy covariance data is not a robust means to obtain V_{cmax25} (the fitted value depends on the model and what other parameters are used in the fitting).

The LSM input parameters do depend on the model setup. We had this highlighted in the main text. Also, we added more discussion on why fitted V_{cmax25} was lower when we fitted g_1 . See related changes in

- Lines 382-385: “Given the under-performance of empirical models when we used a different tuning factor algorithm (on photosynthetic capacity), we highlight it here that (1) inverted model parameters to use in LSMs vary with the model used to fit these parameters, and (2) using parameters inverted from one model setup in another model would likely result in biases in model outputs.”

7. Line 332-336. I was expecting a more thorough discussion of parameter estimation for land surface models. Yes, using prescribed PFT-dependent parameters has shortcomings. Yes, the optimization model may perform better than empirical stomatal models. But the authors have not adequately outlined a strategy for parameter estimation. What I see from their results is that one can estimate parameters by fitting a model to flux tower measurements, but a common parameter (V_{cmax25}) depends on the specific stomatal model and what other parameters are also fitted. I would like to see the authors discuss this further.

This is indeed an interesting point that is worth more explanation. We add a few sentences to clarify that fitted parameters depend on the model used to drive stomatal responses to the environment, and that fitted V_{cmax25} also depends on the stomatal model parameters.

- Lines 367-375 “It is also worth noting that when g_1 was fitted for the empirical stomatal models, our fitting g_1 was higher than CLM defaults, and fitted V_{cmax25} was also closer to OSM (Table 3). The changes in fitted V_{cmax25} was likely due to the higher stomatal conductance caused by higher g_1 (as the model predicted water fluxes increased). For example, if fitted V_{cmax25} did not change when g_1 was higher, then the empirical models would predict higher stomatal conductance, and thus higher photosynthetic rate. In this case, the error between model predicted carbon fluxes vs. flux tower observations would increase. As the BBM and MED predicted carbon flux already centered along the 1:1 line vs. flux tower observations (as in Figs. 12a,b and 13a,b), an unchanged V_{cmax25} would result in higher biases in carbon flux, harming the overall empirical model performance. Therefore, the fitted V_{cmax25} decreased whereas g_1 increased to minimize the error between model predictions and observations.”

Minor

1. Line 5-8. I agree with the comments about big-leaf radiative transfer being unable to resolve vertical gradients in the canopy. However, the current study does not address the impact of vertical gradients in the canopy (though the radiative transfer model does provide the necessary vertical profiles).

The CliMA Land model is designed to be highly modularized, and thus it needs to account for complicated scenarios as well as simplified cases (such as vertical profiles). Being able to account for complicated and more realistic biophysical processes should be a key feature for future LSMs. While the CliMA Land model is able to use a complex setup, we did not have enough data to drive and validate the simulation, and thus we only noted the capability of the model. We have now clarified this in the text the difference and advocated future research to test it:

- Lines 121-125 “We note that our modeling framework allowed us to customize vertical leaf area distribution, leaf angular distribution, and photosynthetic capacity profile vertically. Future research efforts to resolve these distributions within the canopy would make LSMs more realistic in terms of up-scaling of carbon, water, and energy fluxes. Yet, for now we used even distributions in our model simulations due to the lack of knowledge of the true distributions at the study sites.”

2. Line 50-53: These sentences need to be explained better. Why is it impractical to compare models with eddy covariance data? This has been done many times and is a common way to evaluate land surface models.

There is a misunderstanding here. We meant that it is impractical to apply stomatal optimization models at large spatial levels because of the limited number of traits at the site level. However, eddy flux tower data provide good datasets that we can use to invert these unknown bulk traits for stomatal optimization models, and this is what we did in the paper. We have revised the sentence to be clearer:

- Lines 50-51: “While traits used in stomatal optimization models improve predictive skill, the number of traits required to parameterize these process- and trait-based models makes it impractical to apply them at large spatial scales.”

3. Line 58: Why is it ideal to invert required parameters from flux tower observations? Doing so just fits a model to the observations without testing the theory. In fact, this paper demonstrates that. Different estimates for V_{cmax25} are obtained between the BBM, MED, and OSM models. V_{cmax25} is no longer a quantity that relates to leaf gas exchange measurements. It is merely a tunable knob.

We were not saying the traits inverted were ideal. We were saying that these unknown traits can be hopefully inverted from flux tower data, which is an ideal scenario for researchers. To avoid misunderstanding, we have revised the sentence to

- Lines 58-59 “If a high quality flux tower data is used (such as a full suite of environmental conditions and carbon and water fluxes), the traits required to run stomatal optimization models can be inverted from flux tower observations.”

4. Line 67-73. This thought needs to be stated more clearly. What is meant by “simple” and “complex”? Is the point that big-leaf radiative transfer is inadequate and that multilayer models are needed?

Yes, the big leaf model is not adequate to use with remotely sensed data. We referred to the big-leaf model as a simple model and multi-layered model as a complex model given not only the canopy radiation model complexity but also the soil-plant-air continuum that matches the

complexity of the multiple canopy layers. We have revised the sentence now:

- Lines 70-77 “The single leaf representation of the canopy, however, is not adequate in modern LSMs in terms of simulating the reflectance or fluorescence of the entire canopy, which requires bidirectional radiation within the canopy to be simulated. More complex models with multiple canopy layers, horizontal canopy heterogeneity (Braghiere et al., 2021), and more detailed representations of the canopy RT scheme are therefore required for the purpose of simulating canopy optical parameters, such as the RT scheme used in the Soil-Canopy Observation of Photosynthesis and Energy fluxes model (SCOPE; Yang et al., 2017). This way, the advantages of stomatal optimization theory and those of a complicated multi-layered canopy RT scheme are integrated, being able to better relate plant physiology to remotely sensed canopy spectra.”

5. Tables 1 and 2. It is not stated how the data are used. How are chlorophyll, tree density, and basal area used in the model, or are these merely to show that the sites differ in stand structure and physiology?

Thanks for pointing this out. We have revised Tables 1 and 2 to describe how each parameter is used in our model. Please see our revised Table 1 for the changes (pasted below).

Table 1. Site and plant information of Niwot Ridge flux tower site.

Variable	Description	Reference
Site name	Niwot Ridge, AmeriFlux core site US-NR1	
Latitude	40.03°N. Latitude impacts the solar zenith angle, and thus canopy radiation simulations.	Monson et al. (2002)
Longitude	105.55°W	Monson et al. (2002)
Elevation	Height above sea level, 3050 m	Monson et al. (2002)
Canopy height	Canopy height, 12–13 m. A mean canopy height of 12.5 m was used in the model. As to the tree geometry, we assumed the trunk has a height of 6 m, and the canopy spanned from 6 to 12.5 m. We divided the canopy to 13 layers (0.5 m in height per layer). Canopy height causes gravitation pressure drop when computing xylem water pressure profile.	Bowling et al. (2018)
LAI	Leaf area index, 3.8–4.2. A mean LAI of 4.0 was used in the model. LAI affects canopy radiative transfer, and carbon and water flux aggregation.	Monson et al. (2002)
Chlorophyll	Leaf chlorophyll content, 524 $\mu\text{mol m}^{-2}$. Chlorophyll content impact leaf reflectance, transmittance, and fluorescence emission.	Zarter et al. (2006)
Tree density	Trees per ground area, 4000 ha^{-1} . <i>Abies lasiocarpa</i> : 16 trees per 100 m^2 ; <i>Picea engelmannii</i> : 10 trees per 100 m^2 ; <i>Pinus contorta</i> : 9 trees per 100 m^2 . Addressed by basal area per ground area (namely basal area index). Tree density is used to normalize whole plant hydraulic conductance.	Bowling et al. (2018)
Weibull B/C	<i>A. lasiocarpa</i> : $B = 4.28 \text{ MPa}$, $C = 1.47$; <i>P. engelmannii</i> : $B = 4 \text{ MPa}$, $C = 12$; <i>P. contorta</i> : $B = 4 \text{ MPa}$, $C = 4$. Mean $B = 4.09 \text{ MPa}$ and $C = 5.82$ were used. Weibull B/C impacts the tree's water supply capability as well as resistance to drought induced xylem cavitation.	Tai et al. (2019); Choat et al. (2012)
Basal area	Mean basal area per tree. <i>A. lasiocarpa</i> : 0.063 m^2 ; <i>P. engelmannii</i> : 0.08 m^2 ; <i>P. contorta</i> : 0.144 m^2 . Total basal area per ground area for the three species are 0.031 $\text{m}^2 \text{ m}^{-2}$; and thus a mean ground area per basal area of 32.09 $\text{m}^2 \text{ m}^{-2}$ was used in the model. Basal area is used to normalize whole plant hydraulic conductance.	Sproull (2014)
Clumping index	MODIS clumping index, 0.48. A constant CI was used in the test site because of the lack of knowledge on how CI varies with solar zenith angle in the test site. Clumping index impacts the canopy radiative transfer and leaf level light conditions.	He et al. (2012)
Root depth	Root depth, 0.4–1.0 m. A maximal root depth of 1 m was used. Yet, as we prescribed soil water content, the root depth was only used to calculate gravitational pressure drop in the roots. Root depth causes gravitation pressure drop when computing xylem water pressure profile.	Monson et al. (2002)
Soil type	Soil texture class, Cambisol. See Mello et al. (2005) for the detailed van Genuchten parameter for Cambisol type soil. Soil type is used to convert soil moisture to soil water potential.	https://soilgrids.org/
Stomatal model	Ball et al. (1987) model: $g_1 = 9$; Medlyn et al. (2011) model: $g_1 = 2.35 \sqrt{\text{kPa}}$. These empirical parameters are used to simulate stomatal responses to the environment.	De Kauwe et al. (2015)

Yujie Wang (on behalf of all authors)

RC2

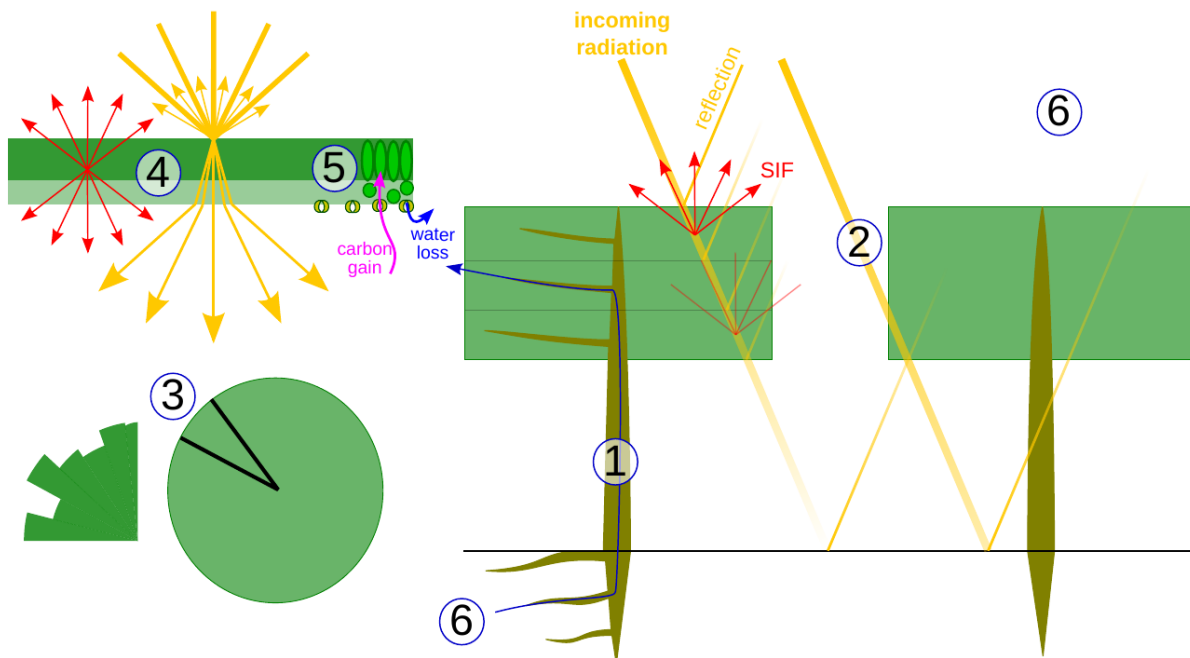
In the study “Testing stomatal models at stand level in deciduous angiosperm and evergreen gymnosperm forests using CliMA Land (v0.1)” by Wang et al., the authors implement and compare three different stomatal models within the CliMA land model. While one model is optimization-based, the other two are empirical. The comparison at two flux-tower locations shows that all models predict site-level carbon fluxes well, while the optimization-based model performs best for water fluxes. Vegetation and Earth system models could benefit from the results of this study by implementing stomatal optimization models. Because of the importance of stomatal conductance for vegetation-climate feedbacks in Earth system models, this paper is an important contribution to the community. The paper is interesting to read and mostly well-written. There are, however, a few unclear passages and the structure of the method Section could be improved. My individual comments are the following:

Dear Reviewer,

Thanks for your appreciation and valuable suggestions. We have revised our manuscript carefully based on the reviewers’ comments. Please see our detailed responses for the changes we have made.

1. An overview figure about the various applied model components and their main inputs would greatly support the understanding of the model framework.

We have now included a diagram showing the important modules or components in the CliMA Land model (as a result, the figure numbering changed). We also attach a copy here:



- 1. Hydraulic traits** such as vulnerability curve and maximum conductance impact water transport, and thus stomatal behavior.
- 2. Canopy traits** such as leaf area index and clumping index impact light penetration to lower canopy, and reflected light and solar-induced chlorophyll fluorescence (SIF) escaping from lower canopy.
- 3. Leaf angular distributions** impact light scattering within the canopy.
- 4. Leaf biophysical traits** such as chlorophyll and carotenoid contents impact leaf level reflectance, transmittance, and SIF spectra.
- 5. Leaf physiological traits** such as maximum carboxylation rate impact leaf gas exchange.
- 6. Environmental conditions** such as soil moisture and atmospheric humidity impact plant's physiological responses.

2. The model description could be generally a bit more streamlined. For example, the authors could give some more background information about used model compartments (e.g. mSCOPE) and describe their changes to the models (especially the land model) in comparison to the original implementation (which is sometimes done, but not always, see minor comments).

We have now included more details to clarify what is new and what is experiment. Since there are quite a few changes, we won't paste all of them below this item. Please find the changes and our responses in the minor comments section.

3. The authors suggest that the optimization-based model could be used for land models within Earth system models. Can the method be applied globally? In L58-66 the problem of global inputs is discussed with a possible solution but from this paragraph, it was not entirely clear to me. It is one of the main motivations of this paper that this method previously was not used in models of larger scale due to a too large number of traits. The paper potentially aims to solve this problem at least partially. It would be nice if the authors could come back to this point in the Discussion.

Using the flux tower to invert global traits maps is a future goal. We have now included a paragraph in the discussion to hint the promise of this approach:

- Lines 390-396 (revised text, clean version; hereafter) “We also highlight it here that using flux tower data to invert site-level bulk traits to use with stomatal optimization has great potential in advancing future land surface modeling. We foresee how global flux tower data could be used to estimate the missing traits, particularly the hydraulic traits. Furthermore, machine learning based algorithms along with climatological data would help solve the issue of sparsely distributed towers. Knowing how these traits vary globally not only helps global simulations using stomatal optimization theory, but also provides a direct way to assess plants' hydraulic health status, helping predict the endangered zones to drought induced tree mortality and potentially shifting traits due to climate change.”

4. In relation to my previous comment, I was wondering why the authors only compared the three models for 2 sites. To evaluate whether the optimization based model could benefit global Earth system modeling, results from different climate zones are required, because stomatal processes could have different properties there. It would be good, if the authors could give a reason, why they only choose 2 flux tower sites and/or why this is sufficient to judge the method on a larger scale.

We thank the reviewer for bringing this potential research up. The most difficult part for this study is to find site-specific traits such as xylem vulnerability curve, so we only choose two representative sites in the present study. Also, as we aim to compare our model simulation to TROPOMI SIF (available from 2018), there are not many sites we can choose from as most flux tower sites do not cover years after 2018. The CliMA Land is a new model, so starting out from a few well characterized sites is the best approach for guiding model development and learning how to run research on more sites. We are currently working on testing CliMA Land model with more flux tower sites with different biome types, and testing more stomatal model representations (we have at least 9 stomatal model alternatives in CliMA Land). Yet, before that, it is useful to verify our new CliMA Land model makes sense to the land model community and is functioning well given the implementation of SCOPE RT scheme along with stomatal optimization models.

5. The 4th. Chapter, Solar-induced chlorophyll fluorescence, does not state exactly which model setup has been used here. Before, three different setups have been compared, but Chapter 4 only uses one of the variants? Generally the aim of this Chapter could be explained a bit more. The authors could also think about restructuring the Chapters to one method Chapter (model description, site description, model protocol etc.) and one Chapter for all the results (comparison of the models and SIF).

Thanks for the suggestions regarding the paper structure. Given the amount of information in the paper, combining all the methods in one section may not be the best option. Thus, we prefer to describe the carbon and water fluxes in one section, and SIF fluxes in another. We made the following changes regarding structure to improve the reading experience (descriptions as well, but too long to paste, please find them in the responses to the minor comments)

- Renamed the 3rd section to “Model evaluation: Carbon and water fluxes”
- Renamed the 4th section to “Model evaluation: Solar-induced chlorophyll fluorescence”
- Restructure each of section 3 and 4 to three subsections:
 - Study sites
 - Model simulations
 - Model performance

Minor comments:

1. L67: Please give a short description, why stomatal optimization models need a RT scheme (maybe add this to a possible overview figure).

Stomatal models work with either simple or complex RT models. However, the simple RT models are inadequate to output canopy reflectance and fluorescence spectra, making it unable to link to remotely sensed data. We have revised the paragraph to clarify this point:

- Lines 70-77 “The single leaf representation of the canopy, however, is not adequate in modern LSMs in terms of simulating the reflectance or fluorescence of the entire canopy, which requires bidirectional radiation within the canopy to be simulated. More complex models with multiple canopy layers, horizontal canopy heterogeneity (Braghiere et al., 2021), and more detailed representations of the canopy RT scheme are therefore required for the purpose of simulating canopy optical parameters, such as the RT scheme used in the Soil-Canopy Observation of Photosynthesis and Energy fluxes model (SCOPE; Yang et al., 2017). This way, the advantages of stomatal optimization theory and those of a complicated multi-layered canopy RT scheme are integrated, being able to better relate plant physiology to remotely sensed canopy spectra.”

2. L74/75: Please state here again in which model the stomatal optimization model and the mSCOPE RT concept are implemented.

Thanks for pointing it out, the sentence was revised now:

- Lines 77-79 “Here, we aim to advance land surface modeling by incorporating a recently developed stomatal optimization model (Wang et al., 2020) and the SCOPE RT concept in the land system of a new generation of Earth System Model within the Climate Modeling Alliance (CliMA).”

3. L74-78: Please also describe here the comparison of the different models, since it is a key result of this paper.

Done. See our revised text:

- Lines 80-82 “We evaluated our model by comparing the model predicted ecosystem carbon and water fluxes to flux tower measurements as well as two well established empirical stomatal models, and the model predicted SIF to TROPospheric Monitoring Instrument (TROPOMI) SIF retrievals (Köhler et al., 2018).”

4. L82: Could the canopy radiative transfer be also used instead of mSCOPE? Or is this not possible for the reasons stated in the introduction?

mSCOPE is not merely a radiative transfer model, it includes RT scheme as well as an empirical stomatal model and energy balance. We only used and adapted the mSCOPE RT scheme in the CliMA Land model (such as canopy clumping and carotenoid absorption). We also included a number of stomatal model alternatives. This is why we were implicitly saying what processes CliMA land model addresses.

5. L87: Please replace “here” by the model name.

We have clarified this in the text:

- Lines 96-97 “In the CliMA Land (v0.1), we treated a site as a uniform “mono-species” stand. Therefore, a suite of average plant traits were applied to the stand, and the stand level simulation was done using these bulk traits.”

6. L87-97: Are these the changes implemented in the land model of CliMA? Which of these points has been part of the model before, and what was newly implemented? What is the difference to the original implementation?

To our knowledge, the hydraulic system in CliMA Land is the most comprehensive and modular among all LSMs. Most LSMs use a simple one-element xylem to simulate plant hydraulics. We have now clarified this advance

- Lines 97-107 “The CliMA Land simulates plant hydraulics numerically using the most comprehensive and modular plant hydraulic system to date. The average plant was represented as a tree, and the modeled tree consisted of a multi-layer root system, a trunk, and a multi-layer canopy to match the soil and canopy setups (Fig. 2a). Each root layer corresponds to a horizontal soil layer, and contains a rhizosphere component and a root xylem in series (water flows through the rhizosphere and then the root xylem). All root layers are in parallel and connected to the base of the trunk. Each canopy layer corresponds to a horizontal air layer, containing a stem and leaves in series (water flows through the stem and then the leaves). All canopy layers are in parallel and connected to the top of the trunk. By default, we accounted for gravity in root and stem (gravity not accounted for in leaves), and thus each canopy layer has its own gravitational pressure drop. Yet, the gravity correction can be customized by setting the height changes of each root and stem. We note here that the hydraulic architecture in the CliMA Land can be freely customized from a single xylem organ to a whole plant with any finite number of root and canopy layers.”

7. 98-104: Also here, please better highlight the actual changes, compared to the original model.

These are original (no model we know in a LSM that simulates plant hydraulics in such details, like xylem water pressure profile can be simulated for each individual leaf).

8. L114: Here the changes to mSCOPE are better introduced, but it would be nice to know how exactly the carotenoid light absorption was implemented in the model.

As requested, now we added an equation describing how carotenoid absorption as APAR is addressed. Changes related to this comments include

- Lines 130-136 “In brief, the relative absorption that is counted as APAR in SCOPE and CliMA Land (kAPAR,SCOPE and kAPAR,CLIMA , respectively) differ in that...

$$k_{\text{APAR,SCOPE}} = \frac{\alpha_{\text{cab}} \cdot C_{\text{cab}}}{\sum (\alpha_i C_i)} \quad (1)$$

$$k_{\text{APAR,CLIMA}} = \frac{\alpha_{\text{cab}} \cdot C_{\text{cab}} + \alpha_{\text{car}} \cdot C_{\text{car}}}{\sum (\alpha_i C_i)}, \quad (2)$$

where α_i is the feature absorption coefficient of trace ingredient (cab for chlorophyll a+b, car for carotenoid), C_i is the content of each ingredient, and $\sum (\alpha_i C_i)$ is the sum of all ingredients (chlorophyll, violaxanthin and zeaxanthin carotenoid, brown pigment, anthocyanin, water, and dry mass). When accounting for carotenoids, APAR-related absorption relative to the total

9. L129: The SLUSPECT-B model should be better introduced. Why is this procedure necessary?

Fluspect-B model is based on PROSPECT model, but outputs fluorescence matrices to compute fluorescence. As we only made one modification in terms of carotenoid absorption as APAR, we do not think it necessary to rewrite all the equations here. However, we highlighted in the revision that FLUSPECT-B is based on PROSPECT and that the inclusion of carotenoid absorption as APAR does not impact the absorption or transmittance, but fluorescence conversion matrices are impacted:

- Lines 149-151 “In the model simulations, we (1) calibrated the leaf chlorophyll fluorescence, reflectance, and transmittance spectra using the FLUSPECT-B model (Vilfan et al., 2016), which advances the PROSPECT model by computing the fluorescence matrices (Jacquemoud and Baret, 1990; Jacquemoud et al., 2009);”
- Lines 157-159 “We note here that as we include carotenoid absorption as APAR, leaf forward and backward fluorescence conversion matrices calculated using FLUSPECT-B model differ from those in SCOPE; however, leaf reflectance and transmittance spectra are the same as in SCOPE.”

10. L160: Why was E_crit defined as the transpiration rate at which leaf xylem hydraulic conductance decreases to 0.1% of the minimum value? Is there a literature example or any clear reason for doing this?

Xylem hydraulic conductance decreases with more negative xylem water pressure, as a result, xylem flow rate cannot be infinite. In the model, we typically set a minimum xylem pressure, and compute the flow rate driven by this minimum xylem pressure, and use it as critical flow rate (xylem flow rate cannot physically go beyond this value). Sperry and Love (2015) and Sperry et al. (2016) have some classic plots of the xylem water supply curve indicating this maximum flow rate, and we have now included these citations. Also, we had the supply curve plotted in Figure 4a: when Psi gets more and more negative, xylem flow rate actually saturates, and the saturating flow rate is Ecrit. We also showed in Figure 4b that Ecrit decreases when soil water potential gets more negative. Changes related to this comment include

- Lines 176-178 “and Ecrit is the critical transpiration rate for that leaf in mol m⁻² s⁻¹, beyond which the leaf hydraulic conductance drops to 0.1% of the maximum value (0.05% in Sperry and Love, 2015; Sperry et al., 2016).”

11. L166/167: Please revise sentence structure.

We have revised the sentence now:

- Lines 190-194 “Note that the risk term in equation 5 has the same mathematical form as equation 11a in Wang et al. (2020), but the two differ in that equation 5 uses leaf-level flow rates so as to use with our adapted SCOPE RT model, whereas equation 11a in Wang et al. (2020) model uses mean canopy flow rates to use with the big leaf model. Therefore, Ecrit in the CLIMA Land differs among canopy layers given the different gravitational pressure drop and xylem pressure profiles.”

12. L208: Could soil moisture, leaf temperature and LAI be modeled by the land model?

Yes. We are working on a manuscript to test the soil model and its coupling to the vegetation. As canopy energy balance also depends on soil energy budget (long wave radiation and soil reflectance), we will present these developments step by step. Leaf area index is used as an input in this study. However, when we reach the stage to use CliMA Land to predict future plants' responses to future climate, we will need to optimize LAI.

13. L209: Please describe the "offline simulations" a bit more. E.g. how many simulations under which conditions.

Offline simulations mean that the water and carbon fluxes do not impact the environmental conditions since we used prescribed conditions. We have clarified this in the revision

- Lines 242-243 that "and then we ran offline simulations (namely carbon, water, and energy fluxes do not feedback to the environmental conditions)."

14. L214: What uncertainty is reduced here? Of the evaluation data or of the model?

Thanks for pointing it out. We meant to reduce the uncertainty in evaluation data. As pointed out by reviewer 1, it is possible and normal for researchers to compare model predictions to GPP. However, GPP is a product that is derived from NEE. Thus, we preferred to compare model output directly to NEE, and this is the uncertainty we try to reduce here. We have clarified this in the revision:

- Lines 248-250: "To further reduce the uncertainty in evaluation flux tower data when comparing model simulations to observations, we compared the modeled carbon and water fluxes directly to flux tower estimations rather than reprocessed products such as gross primary productivity (GPP)."

15. L225-234: It is not clear for several of these points, which steps are newly implemented by the authors and which were already part of the land model.

The CliMA Land model is a new land surface model, so all the components from lines 225 to 234 are new in the CliMA Land model. We note here that only section 2 described the new implementations compared to other LSMs, and section 3 described the experiments we did using the CliMA Land model. To clarify these, we have now renamed sections 3 and 4 to denote those sections were applications of the CliMA Land model.

- Section 3 to "Model evaluation: Carbon and water fluxes"
- Section 4 to "Model evaluation: Solar-induced chlorophyll fluorescence"

16. L245: Is the root:stem:leaf resistance ratio of 2:1:1 a widely accepted ratio? Is there a literature example for this choice?

This ratio was the default set up of the Sperry hydraulic model. We had conversations about this ratio when we developed the Sperry stomatal model in 2016-2017, and Dr. John Sperry came out with this ratio by looking into the published xylem pressure profiles: root xylem pressure (estimated using soil water potential), midday tree based xylem pressure, midday stem xylem pressure, and midday leaf xylem pressure (often known as leaf water potential). Pressure drop from soil to tree base approximately accounts for 50% of the total pressure, and stem and leaf pressure drop account for about 25% each. I also had the pressure profile measured as well in Wang et al. (2019) The stomatal response to rising CO₂ concentration and drought is predicted by a hydraulic trait-based optimization model, and the ratio was also approximately 2:1:1. However, this ratio is not widely confirmed at many sites or biomes. Future research into the details of hydraulic segmentation will improve the understanding. We have now clarified in the text of the source for this ratio

- Lines 280-281 “(we assumed a constant root:stem:leaf resistance ratio of 2:1:1, consistent to the ratio used by Sperry et al. (2017))”

17. L249: thoughtout -> throughout

Thanks for pointing out this typo, and we have fixed it.

18. L304: Here would be a good place to discuss the problems to implement stomatal optimization in vegetation models. The authors stated before, that this is problematic due to missing trait data. How exactly does the described approach help here and what is missing for a global application possibly in Earth system models?

Thanks for pointing it out, and this is really a good point. We have added a new paragraph in section Land model parameterization.

- Lines 390-396 “We also emphasize that using flux tower data to invert site-level bulk traits to use with stomatal optimization has great potential in advancing future land surface modeling. We foresee how global flux tower data could be used to estimate the missing traits, particularly hydraulic traits. Furthermore, machine learning based algorithms along with climatological data would help solve the issue of sparsely distributed towers. Knowing how these traits vary globally not only helps global simulations using stomatal optimization theory, but also provides a direct way to assess plants’ hydraulic health status, helping predict the endangered zones to drought induced tree mortality and potentially shifting traits due to climate change.”

19. L315: This sounds as if the underperformance of the empirical models were based on an arbitrary decision of the authors. Please rephrase.

Thanks for your support. And this is not only our arbitrary decision, other LSMs may also suffer from this imperfect parameterization. We have now revised the sentence to clarify this in the main text:

- Lines 354-358 “This under-performance may result from imperfect land model parameterization, which was adopted in our model simulations. For example, CLM uses a constant g_1 for a plant functional type regardless of where the plant grows (in a wet or dry region); also, g_1 is estimated using gas exchange measurements for well watered plants, and thus may not well represent the scenario of drought stress. Furthermore, the use of a V_{cmax25} tuning factor interfered with the prescribed g_1 .”

20. L325/327: Rephrase; the alternation of g_1 *within the empirical models* shows potential...

Thanks. It is now revised to

- Lines 376-378 “The alteration of g_1 within the empirical models shows potential in better capturing carbon and water fluxes than the tested stomatal optimization model (Table 4); and we believe more site-specific g_1 setups would improve the empirical model predictive skills.”

21. L322-331: A good solution for this problem would be to do model simulations at other sites or even potentially global. A short reason, why this has not been done in this paper, would be good here.

We have now added a sentence to highlight it:

- Lines 387-389 “As such, we recommend to revisit and re-calibrate the land model parameterization based on the stomatal model and tuning factor algorithm that was used for each LSM based on real measurements.” to suggest more cautious model parameterization.

- Lines 390-396 “We also highlight it here that using flux tower data to invert site-level bulk traits to use with stomatal optimization has great potential in advancing future land surface modeling. We foresee how global flux tower data could be used to estimate the missing traits, particularly the hydraulic traits. Furthermore, machine learning based algorithms along with climatological data would help solve the issue of sparsely distributed towers. Knowing how these traits vary globally not only helps global simulations using stomatal optimization theory, but also provides a direct way to assess plants’ hydraulic health status, helping predict the endangered zones to drought induced tree mortality and potentially shifting traits due to climate change.”

Yujie Wang (on behalf of all authors)

CC1

In this study, Wang et al. present an interesting analysis where they evaluate one stomatal optimisation and two empirical stomatal models in the new CliMA model. In contrast to many land surface schemes, this model is able to account for greater complexity in the treatment of the canopy space, hence this paper has the potential to offer interesting insights to the state of the knowledge, in particular as we think about more realistic scaling up (from the leaf) of water fluxes. Nevertheless, I currently have some questions about the presentation of the methodology that I think are worth clarifying for the reader, I will outline these below. (Note I've not read beyond the methods...)

Dear Dr. De Kauwe,

Thanks for the constructive suggestions and comments, and they were really good points. We have now covered these points in our revision. Please find our responses and changes per item below.

- One overall concern I have is about an apparent "conflict" in added complexity in some assumptions with marked simplifications in others. Are these tradeoffs warranted(?), it would be good to add some commentary on this point. For example, a lot is made of the vertical treatment of CliMA but then you assume constant leaf physiology parameters throughout the canopy - so V_{cmax} does not change with depth through the canopy? I make additional comments below.

As the CliMA Land model is more complex than most existing land surface models in its implementation of plant physiology, it is fairly difficult to constrain all model inputs at this early stage of model development. For example, we do not quantitatively know how leaf phenology and physiology differs among canopy layers spatially or temporarily, such as V_{cmax} . Yet, the advantage is that the CliMA Land is highly modularized and flexible. Even though these complexities were not tested in the current manuscript, it is important to highlight the capability of the model so as to promote future model evaluations once we have enough data to constrain the model. Regarding vertical V_{cmax} in particular, we have a research paper in review that compares the scenarios with and without a vertical V_{cmax} profile (see <https://doi.org/10.5194/bg-2021-214>). We have now included the following changes in the main text to clarify this and to advocate future research efforts:

- Lines 90-94 (clean version, and hereafter): "We note here that, compared to most land surface models (LSMs), we implemented more complex biophysics in the CliMA Land, such as hyperspectral canopy radiative transfer scheme and multi-layer canopy hydraulics. These detailed features, along with the high modularity of CliMA Land (such as turning on and off detailed features), allows users to perform research with different complexities and at multiple levels from leaf to global scales (e.g., Wang and Frankenberg, 2021)."
- Lines 122-124: "Future research efforts to resolve these distributions within the canopy would make LSMs more realistic in terms of up-scaling of carbon, water, and energy fluxes."

- In your parameterisation of the stomatal models (table 1), you used species-level hydraulic parameters to determine the Weibull function in the optimisation model + fit K_{max} to site data, but by contrast, you used plant functional type parameters to run the empirical models. Is there any evidence these values are appropriate for the species at these two flux sites? Isn't this akin to "calibrating" the optimisation model but then evaluating its skill improvement

relative to an uncalibrated (empirical) model? I also note that the correct citation for the Meldyn g_1 values is De Kauwe et al. 2015, GMD not the CLM tech note (as this borrows from that paper and isn't the original source).

As to the trait inputs for the models, we used the species-level traits as estimates of the bulk property of the stand, and the identical suite of traits was used for the tested stomatal optimality model and two empirical stomatal models. To be more specific, the xylem vulnerability curves were also used to compute the leaf-level hydraulic conductance for the empirical models using exactly the same hydraulic model, and then used to calculate the tuning factor for the empirical stomatal models. We considered it a fair comparison because (1) the same number of fitting parameters were used for all the tested models, and (2) the fitted K_{max} not only affects the stomatal optimization model but also impacts the empirical stomatal models in that the stress tuning factor was parameterized as K/K_{max} . Further, we also considered the case of what if the unsatisfactory performance came from the g_1 parameter we used, and ran an extra fitting for g_1 for the empirical stomatal models. Changes related to this comment include:

- Lines 364-370, you can find that when we fitted g_1 for the empirical models, the model performance improved (current Figs. 13 and 14, Table 4).
- Pages 14-15, thanks for pointing out the original source for the g_1 parameters, and now we have the reference updated as in Tables 1 and 2 and reference section.

- Later when you do fit values (Table 3), are these values sensible? There must be some site estimates of V_{cmax} that the values could be compared to? I note that the Niwot values are pretty low from a cursory glance. Similarly, a g_1 of 16 makes no sense to me, if you look at Lin et al. 2015 NCC, it is above any of the values they derive in their global synthesis. I think these fits are worth double-checking. Similarly, can you also double-check the K_{max} values?

These fitted parameters are bulk traits for the site, and thus likely differ from leaf-level observations, though should be of the same magnitude as measurements. Regarding a fitted g_1 of 16 that was way higher than CLM default and those in Lin et al. (2015), this was due to the way the stress tuning factor was applied. Since the tuning factor was applied to V_{cmax} , g_1 needs to be higher to account for the changes in V_{cmax} compared to the scenario of tuning g_1 (typically done in LSMs), which is discussed in the section "Land model parameterization". The V_{cmax} was low in Niwot Ridge as it is an evergreen needle-leaf forest. As we did not have all the V_{cmax} for the species at Niwot Ridge, we were not able to say in the text that the fitted mean V_{cmax} was in a realistic range. However, according to Timothy Tomaszewski & Herman Sievering (2007), V_{cmax} for spruce ranged from 8-12 and J_{max} ranged from 46-57. A V_{cmax} estimate of 15 for OSM seems to be reasonable. Yet, for MED and BBM, V_{cmax} is higher because of the tuning factor. As to K_{max} , we did not have measurements to back it up. Also, for fitted g_1 , we added a sentence in the text to highlight that g_1 parameters are probably highly sensitive to the model setup. Associated changes to the text include:

- Lines 382-385 (Reviewer 1 pointed it out as well): "Given the under-performance of empirical models when we used a different tuning factor algorithm (on photosynthetic capacity), we highlight it here that (1) inverted model parameters to use in LSMs vary with the model used to fit these parameters, and (2) using parameters inverted from one model setup in another model would likely result in biases in model outputs."
- Lines 367-375 "It is also worth noting that when g_1 was fitted for the empirical stomatal models, our fitting g_1 was higher than CLM defaults, and fitted V_{cmax25} was also closer to OSM (Table 3). The changes in fitted V_{cmax25} was likely due to the higher stomatal conductance caused by higher g_1 (as the model predicted water fluxes increased). For example, if fitted V_{cmax25} did not change when g_1 was higher, then the empirical models would predict higher stomatal conductance, and thus higher photosynthetic rate. In this case, the error between model predicted carbon fluxes vs. flux tower observations

would increase. As the BBM and MED predicted carbon flux already centered along the 1:1 line vs. flux tower observations (as in Figs. 12a,b and 13a,b), an unchanged V_{cmax25} would result in higher biases in carbon flux, harming the overall empirical model performance. Therefore, the fitted V_{cmax25} decreased whereas g_1 increased to minimize the error between model predictions and observations.”

- Lines 304-309 “Given that the fitted parameters were bulk properties of the sites, we expected them to differ from leaf-level observations but be of the same magnitude. However, because of the limited direct measurements in the studies forest sites, we were only able to find one study reporting a V_{cmax25} ranged from 8 to 12 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and a J_{max25} ranged from 46 to 57 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at Niwot Ridge (Tomaszewski and Sievering, 2007). Therefore, the OSM estimated $V_{cmax25} = 15 \mu\text{mol m}^{-2} \text{s}^{-1}$ seemed to be reasonable; and as we explained, BBM and MED estimated V_{cmax25} had to be higher than the OSM estimate due to the tuning factor.”

- Furthermore, why is V_{cmax} being fitted differently across schemes? This is a unique quantity that reflects the canopy, so by varying it across models, aren't we shifting errors between water and carbon fluxes? When g_1 is not fitted at the MOFLUX site, V_{cmax} in the OSM model is half of what it is in the empirical model, this is not a small difference. I note there is an explanation but in the text but this isn't clear to me, "effective V_{cmax} ", I don't see why this would (a) differ across approaches and (b) why it would ever differ by so much.

Typically the beta function (a tuning factor) used to force stomatal response to drought is done on g_1 , but this way involves extra beta function parameters such as a linear or nonlinear curve relating to soil moisture. In this paper, we adopted the approach used in Kennedy et al. (2019) that uses a tuning factor to decrease V_{cmax25} under stressed conditions. Thus, the effective V_{cmax25} used to compute photosynthetic rate is lower than the fitted V_{cmax25} . As a result, the fitted V_{cmax25} should be lower than that from the stomatal optimization model. Since the tuning factor is used on V_{cmax25} , stomatal response to VPD is affected. So, the g_1 has to be higher to make the stomatal responses to VPD correct. This is why our fitted g_1 was much higher than that from the original papers. To make it clearer, we added a beta term in equations 4-6 to highlight that A_{net} for BBM and MED models changes with the tuning factor. Related changes can be found at

$$g_{sw} = g_0 + g_1 \cdot RH \cdot \frac{A_{net}(\beta_w)}{C_s}$$

- Line 197

$$g_{sw} = g_0 + 1.6 \cdot \left(1 + \frac{g_1}{\sqrt{D}}\right) \cdot \frac{A_{net}(\beta_w)}{C_a}$$

- Line 203

- By prescribing leaf temperature and soil moisture we are not able to get a good sense of how this would actually work in a LSM scheme when these feedbacks would be important. While I think this approach has value for minimising differences across models, I think it is equally valuable to turn on these feedbacks. I would prefer to see both versions presented.

We are preparing papers to test the soil water and energy budget module (including water phase changes in soils), and we will show the model performances of soil water and energy budget in those papers. Leaf energy budget in the CliMA model is coupled to soil surface temperature, and soil water temperature profile. It is beyond the scope of the current paper, so we will test how well CliMA Land performs regarding the water and energy budget in the soil and canopy airspace in the future along with a fully functional soil module.

- Is there a citation to support the division of Kmax (speaks further to my point about apparent complexity vs simplification)?

The partitioning of Kmax into different organs allows us to more realistically represent plant hydraulics. The CliMA Land is highly modularized, allowing us to use it with different setups. For example, we can simply model water transport using a tree with only one xylem element like a leaf directly attached to the soil. Yet, by default, we used the more complex and realistic scenarios as we have multiple root and canopy layers. Currently, some LSMs use the big-leaf concept and a single leaf hydraulics component, but this does not allow for a heterogeneous canopy flux simulation.

This partitioning of root:stem:leaf was the default set up of the Sperry hydraulic model. We had conversations about this ratio when we developed the Sperry stomatal model in 2016-2017, and Dr. John Sperry came out with this ratio by looking into the published xylem pressure profiles: root xylem pressure (estimated using soil water potential), midday tree based xylem pressure, midday stem xylem pressure, and midday leaf xylem pressure (often known as leaf water potential). Pressure drop from soil to tree base approximately accounts for 50% of the total pressure, and stem and leaf pressure drop account for about 25% each. I also had the pressure profile measured as well in Wang et al. (2019) The stomatal response to rising CO2 concentration and drought is predicted by a hydraulic trait-based optimization model, and the ratio was also approximately 2:1:1. However, this ratio is not widely confirmed at many sites or biomes. Future research into the details of hydraulic segmentation will improve the understanding. We have now clarified in the text of the source for this ratio

- Lines 280-281 “(we assumed a constant root:stem:leaf resistance ratio of 2:1:1, consistent to the ratio used by Sperry et al. (2017))”
- Lines 90-94 (clean version, and hereafter): “We note here that, compared to most land surface models (LSMs), we implemented more complex biophysics in the CliMA Land, such as hyperspectral canopy radiative transfer scheme and multi-layer canopy hydraulics. These detailed features, along with the high modularity of CliMA Land (such as turning on and off detailed features), allows users to perform research with different complexities and at multiple levels from leaf to global scales (e.g., Wang and Frankenberg, 2021).”
- Lines 105-107 “We note here that the hydraulic architecture in the CliMA Land can be freely customized from a single xylem organ to a whole plant with any finite number of root and canopy layers.”

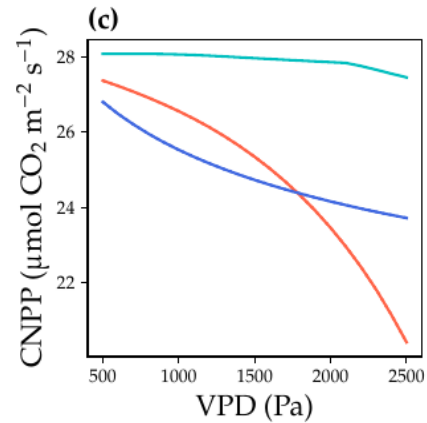
- Should eqn 7 have an autotrophic respiration loss term?

Leaf respiration was accounted for in Anet. Wood respiration was addressed in Rremain as in equation 11.

$$R_{\text{remain}} = R_{\text{base}} \cdot f(T_{\text{soil}}) = R_{\text{base}} \cdot \left(\frac{T_{\text{soil}} - 298.15}{10} \right)^{Q_{10}}$$

- Fig 5c, the axis is Pa not KPa.

Thanks for pointing out this error. We have now fixed it. See our revised Figure 6c (attached below; a new figure 1 was added, so Figure numbering changed)



In summary, I think this will be of great interest to readers as this approach is novel in terms of the role of canopy complexity and stomatal optimisation but there are key methodological points that warrant clarification. I look forward to reading a future revised version.

Thanks for pointing out the original source for the model inputs, typos, and points that needed clarification. We have fixed them in the revision, and believe the model will benefit the LSM community.

Yujie Wang (on behalf of all authors)

References

Timothy Tomaszewski & Herman Sievering (2007) Canopy uptake of atmospheric N deposition at a conifer forest: Part II- response of chlorophyll fluorescence and gas exchange parameters, *Tellus B: Chemical and Physical Meteorology*, 59:3, 493-501, DOI: 10.1111/j.1600-0889.2007.00265.x