

## Referee 1

Ding et al present an interesting study using the FATES model at the Soaproot site in the southern Sierra Nevada Mountains, USA, which is dominated by ponderosa pine trees. The authors explore parameter space for root depth, hydraulic, and stomatal parameters in their experimental design. The authors initialize FATES with the observed demography and turn off growth and mortality to observe how changes in climate over a major drought period impact simulated physiology, soil moisture, and water and carbon fluxes. The model is forced with 4km resolution MACA climate. Model predicted ET and GPP are compared to flux tower observed LH and tower modeled GPP. The exploration of hydraulic parameter space and rooting depths is a really interesting and important set of model experiments to perform for hydraulically enabled vegetation models and their application to terrestrial ecosystem processes. However, I have several major methodological concerns that I hope the authors can address

*We want to thank the reviewer for their positive assessment of the value of the manuscript.*

Method clarity: Some aspects of the experimental design are not clear. For example, how was the soil moisture initialized or spun up? Did the authors test the sensitivity of their conclusions to this method? What is the vertical resolution of the new multi soil layer model? Given that soil water is fundamental to these experiments, I think these are important details. I also wonder why the authors forced the model with MACA instead of the flux tower met.

*We set the initial soil water content to be 75% of saturated water content, close to field capacity. We believe this is the realistic value because the model is initialized in Jan, when the study area has high precipitation and trees are all in a dormant status, and in a year when there is not drought. We found that initializing with very dry soils led to numerical errors in the hydraulics model, but have not tested the sensitivity of the results to this assumption beyond that. We also performed a sensitivity experiment to initialize the model with higher soil water content. There is no obvious effect on the outcome. We incorporated this content in methods section (L393 – 398)*

*The vertical soil is set to be 8 meter deep and divided into layer and the thickness of each layer (meter) is given as the following:*

0.015639592	0.010145736	0.01672749	0.027578969		
0.045470033	0.074967411	0.123600365	0.203782551	0.335980626	0.553938405
0.913290032	1.505760701	2.482579697	4.093081953	6.748351278	11.12615029
18.34392065	30.24401217	49.86394617			

*We use MACA data for two reasons. The main reason is that the flux tower data has long gaps with missing data, and thus needs filling with some reanalysis-based product. Secondly, the data has been used in a previous study for driving CLM (Buotte et al., 2021). We have compared the MACA data with the flux tower data and they are similar. We show this comparison*

*between MACA and flux tower measurements, and show that the MACA data is not unreasonable (see added fig S8).*

Assessment of model performance: I see this as a contextualized OSSE experiment. However, I do think that the authors should do a little more than calculate TMSE relative to GPP/ET. Perhaps use some of the standard ILAMB metrics in addition to RMSE like inter annual variability, monthly variability and phase shifts in annual cycles?

*When have considered whether it is possible to use the ILAMB approach for this, in particular some of the metrics beyond RMSE. However, we note that the datasets in ILAMB are mainly global-scale and thus not appropriate for the local-scale analysis in this study. . Likewise, many of the metrics described in Collier et al. (2018) are most appropriate for large-scale and/or long-term datasets. Thus we believe that the RMSE is the most relevant metric for this analysis. We incorporated above content in methods section (L413 - 416)*

*“We choose RMSE as it is a common and compact metric of assessing model performance, though we note that other metrics could in principle be used, each of which has different advantages and disadvantages (e.g. Collier et al., 2018).”*

The authors come off as defensive about the model predicted leaf water potentials, understandably because they are a physical (there are basically no trees that allow for LWP lower than -4 MPa, the -10 MPa cited in the text is for California chaparral and the correct citation is Tyree 1997 not Vesala 2017). I do appreciate that the experiments are designed to test relative sensitivities of physiological diagnostics to model parameters and that ecosystem dynamics are turned off, but with some of the parameterizations all the trees would be VERY dead before the drought started. I think with such ridiculous LWP values, the authors are going to lose the confidence of a large portion of their audience that has a physiology but not a modeling background, so I hope they will try to make some modifications to their experiments. The fact that the LWPs are dropping so low suggests that the authors might want to reconsider the vulnerability curve parameterization and parameter space that they explore for their experiments. Another option is to use the simulated water potentials to tell us more about the system (for example, to screen what parameter combinations are physiologically impossible at the site). It seems like these trees must have really deep roots to exist at this site with is in agreement with the conclusions of Goulden and Bales 2019. Why not direct the discussion in this direction instead?

*Thanks for pointing this out. Again, we reiterate that if we allow the trees to die in the experiment, then this will fundamentally change the interpretation of the results by making them a function of ecosystem structure and traits, rather than a function of traits alone conditional on the same fixed structure. We feel the second option you suggested matches better with the goal of this study and showing the advantage of using the model. We have incorporated the above content in the discussion section (L634-636). “. Further, the unrealistic leaf water potential*

*from the shallow root simulations indicates that the trees at that site must have really deep roots to exist at this site, whichwith is in agreement with the conclusions of Goulden and Bales (2019).*

”

### **Minor line specific comments**

There are grammatical problems throughout the text which could use further proofreading (tense problems etc)

*We have done a more thorough proofreading.*

Almost everywhere water potential units are written as ‘Mpa’ when they should be ‘MPa’. Similarly, the authors should be consistent with capitalization/abbreviation of ‘Fig.’, ‘fig’, ‘Figure’ throughout the text. Id also like the authors to denote the denominator as either  $1/x$  or  $x-1$  rather than using both in the text

*Thanks for the suggestion. We have made changes accordingly*

Instead of using kLWP as one of the parameters, can the authors choose a different letter, this is easily confused with conductance (k)

*Thanks for the suggestion. We used a different symbol:  $W$*

All of the figures would benefit from increased font size.

*We increased the font size*

L359-361 I am having trouble understanding what the authors mean here, can they clarify?

*We rephrased that sentence for better clarity.*

## Referee 2

This manuscript presents a new model of plant hydraulics in the framework FATES, named FATES-Hydro, and kind of explores its sensitivity on some parameters for the ponderosa pine forest US-CZ2.

The manuscript is a mixed bag of several elements, which leaves the reader alone on most of it. It might be that it presents the model development of FATES-Hydro (but this is not clear), it does a very informal sensitivity analysis, and it assesses model behaviour during a very long drought without the pretence of realism.

*To clarify, this is not a simple model development paper, nor a strict sensitivity paper. The main purpose is to use some sensitivity analysis to explore scientific questions. We made some manipulations on the original hydraulic module that has been written by Christoffersen et al, and that is described in more detail by Xu et al., in review, <https://doi.org/10.5194/egusphere-2023-278> and its supplementary material. We will add this statement in the introduction section to clarify this point and describe the specific changes that have been made relative to those also described in Xu et al. We added above content in the introduction and methods section (L157 – 160 205 - 206) “We note thatNote, this is not an exhaustive strict model parameter sensitivity study. TRather, the main purpose is to use a sensitivity analysis to explore scientific questions around hydraulic trait tradeoffs.” & “FATES-Hydro is described in more detail by Xu et al., (in review, <https://doi.org/10.5194/egusphere-2023-278> ) and its supplementary material.”*

*We acknowledge that the biggest disadvantage of this study is the lack of sufficient field data to constrain the model. This is a result of using a natural drought as an experiment of opportunity, which because it was not anticipated, did not allow for as coordinated planning as would be the case in an experimentally-manipulated drought. The trees at that site had all died by the time we started this study. We made this point in the discussion section (L326-303).*

Was the newly model developed for the study? The text reads (somehow) as if it should have been presented in Fisher et al. (2015) and Koven et al. (2020), but they are rather for FATES itself. So I guess the model development is presented here for the first time. If not, the reference is missing.

*As we note above, the more fundamental reference for the FATES-Hydro model is Xu et al., in review, <https://doi.org/10.5194/egusphere-2023-278>, which had not yet been submitted at the time of this manuscript’s submission. We will rework some of the text to refer to more details referenced in the Xu et al. manuscript, and how the specific configuration here differs from that. (L202 - 206) “FATES-Hydro is a recent development of the FATES model (Fisher et al., 2015; Koven et al., 2020), in which a plant hydro-dynamic module, originally developed by Christoffersen et al. (2016), was coupled to the existing photosynthesis and soil hydraulic modules. FATES-Hydro is described in more detail by Xu et al., (in review, <https://doi.org/10.5194/egusphere-2023-278> ) and its supplementary material.”*

- If yes, it is presented very badly: It changes notation all the time, for example using LWP or  $\Psi_l$  for water leaf potential.  
*We now used LWP for the leaf water potential*
- It uses unusual notation such as  $S_e$  for saturation (called standardized relative water content in the manuscript).  
*We used  $S_e$  because this is the notation used in the original Van Genuchten (1980) paper*
- It uses strange definitions such as " $e_i$  is the saturation vapor pressure (Pa) inside the leaf at a given vegetation temperature when  $A_n=0$ ", which might be true but it is not explained (sounds strange anyway, why saturation should depend on  $A_n$ ).  
*This is a mistake, we removed "when  $A_n=0$ " in the revision*
- There are different parts of the model that are not connected in the manuscript. For example, how is the formulation of Vesala et al. (2017) connected to the rest such as Ball-Berry?  
*The term(formula) of Vesala et al.(2017) is used to downscale saturated inner leaf vapor pressure in the transpiration. This is explained in detail in the FATES tech notes.*
- Also in Vesala et al. (2017), what is  $k_{LWP}$  in the Kelvin equation? It is not given in Vesala et al.  
*This is an additional scaling term we introduced in our version to take into account the deviation of the internal structure of a real leaf from the perfect physical condition in Vesala et al.'s paper.*
- It is not explained how  $\Psi_l$  is calculated.  
*This is explained in Xu et al.'s paper, we provided the reference in the revision. Please see response to the last comment.*
- Are there several stem sections?  
*There is one stem section per cohort.*
- It is not mentioned how root water uptake is calculated. I seem to have guessed at one point in the result section that it might be proportional to root length density.  
*That is correct. We will explain this better in revisions.*
- What is "we have sequentially solved the Richards' equation for each individual soil layer"? How is this working?

*This means we solve the equation for soil layers one by one, not simultaneous solve the equations for all the layers at the same time. There is more detail on this in Xu et al., in review.*

*We have made corrections accordingly for all of the above, and added the detailed description of the hydraulic module in both the FATES tech notes, and in Xu et al. manuscript. See response to the last comments.*

There are also weird choices like using Ball-Berry while calculating leaf water potential. There are a number of good papers that discuss this such as Anderegg et al. (PLoS One 2017, 10.1371/journal.pone.0185481) and references therein.

*To clarify, we didn't employ the Ball-Berry model in calculating leaf water potential, instead we incorporate the constraint of leaf water potential on photosynthetic rate and stomatal conductance through leaf water potential. Leaf water potential is calculated purely based on the physical process of transpiration. There are many different stomatal models used in the literature, which reflects real epistemic uncertainty in how to handle the process of stomatal conductance.. There is no right or wrong stomatal conductance model. Moreover, several studies (e.g. Franks et al. 2018, <https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.14445>; or Knauer et al., 2015, <https://doi.org/10.1002/2015JG003114>) have shown a high degree of overlap between stomatal models such that the parametric uncertainty within any given stomatal model is greater than the structural uncertainty associated with which model to choose. In any case, this uncertainty was not the focus of the paper, so we chose a single stomatal model and used it consistently across scenarios.*

It would also have been interesting to know why the factor  $\beta_t$  (why  $t$ ?) is applied on  $V_{cmax}$  when using actual leaf water potential. What is the physiological mechanism that reduces  $V_{cmax}$  during the day following leaf water potential?

*In the model,  $\beta_t$  is an overall measure of leaf water content. Low water content can reduce the biochemical reaction of the photosynthesis processes as those reaction requires water. In addition, reduced water content may also increase leaf temperature which may further inhibits the photosynthesis. We agree that there is further structural uncertainty in terms of how the hydraulic limitations affect both photosynthetic and conductance processes, and plan to explore this in future work. We discussion this issue in the methods section (L321 - 324) ". We acknowledge there can be structural uncertainty in terms of how the hydraulic limitations affect both photosynthetic and conductance processes, and plan to explore this in future work. "*

However, I was wondering why the model was developed in the first place given that the current manuscript uses CLM5 in FATES, which already includes plant hydraulics.

*The difference is that CLM5 is not a demographic model but a big-leaf model, while FATES is a demographic model that allow us to simulate the vegetation dynamics through demographic processes. Because the fundamental unit shifts in this transition from a PFT to a cohort of a given size, canopy layer and PFT on a given patch of shared disturbance history, this requires a different approach to representing hydraulic dynamics than in the unstructured big-leaf model. We have described this aspect of FATES model in original MS (L197 – 200) "FATES is a cohort-based, size- and age-structured dynamic vegetation model, where long-term plant growth and mortality rates and plant competition emerge as a consequence of physiological processes. "*

There are also different developments of plant hydraulics in the literature such as all the work about the models SurEau and FETCH, Janott et al. (Plant and Soil 2011, 10.1007/s11104-010-0639-0), Huang et al. (New Phytolo 2017, 10.1111/nph.14273), to name just a few that could have been considered or discussed.

*We have discussed these different hydraulic models in the MS, L 212 – 214 , where we indeed cited Janott et al paper, though not Huang et al. We will cite the Huang et al paper in the revised manuscript as well. “The second group represents plant hydraulics by a series of connected porous media, corresponding to each plant compartment (e.g. Bohrer et al. 2005, Janott et al. 2011, Xu et al., 2016, Christoffersen et al., 2016)”*

For me, the really interesting part would have been the interactions between soil water and cohorts. It sounds like that the model has one tile per cohort so there is no interaction between cohorts. This is not explained in the manuscript let alone explored or discussed.

*In the model, multiple cohorts grow on the same land unit, and share the soil water. So there are interactions between cohorts. We will make this point more clear in the revised manuscript. In this case, we wanted to first understand the direct trait control in the absence of structural differences, which is why we held ecosystem structure constant and turned off growth and mortality processes. However, the model is capable to investigate dynamic competitive effects when growth and mortality processes are turned on. This will be the next step that we plan to explore in subsequent work. And this is also the reason why we choose to use FATES-Hydro. We have talked about this in MS (L378-384)*

The manuscript presents further an ad hoc sensitivity analysis. It is also not really explained so I have to guess. Presenting sensitivity as change of model output due to change of a parameter within limits depends, of course, on the chosen limits. "the effective rooting depth, above which 95% of root biomass stays, varies from 1m to 8m", which is such a large range that, of course, everything will depend on it. I need no hydraulic model to know that.

*We choose the range of variation of these parameters based on the allowable biological range. If the range of root distribution can be that large in reality, which has been shown in many studies, then root distribution is the most important plant traits that can affect the response of the trees at that site. We are not trying to find which is the most sensitive parameters for a fixed fractional change in parameter value, but rather to identify what are the most important plant traits in that region. We emphasize this point in the introduction section (L160 - 162). “We note thatNote, this is not an exhaustive strict model parameter sensitivity study. TRather, the main purpose is to use a sensitivity analysis to explore scientific questions around hydraulic trait tradeoffs.”*

*In addition, which traits/parameters have strongest impact on model outcome largely depends on the physical environment and the temporal pattern of their variation (L320 - 325). “Because this design requires a relatively small set of parameters or groups of parameters to vary, we chose parameters that represent the major axes of relatively well- understood stomatal, xylem and rooting mechanisms/strategies that control the hydraulic functioning of trees. We set the values of these parameters within the realistic (allowable biological) range based on online database, and literatures where the species and physical environment are as close to our system as possible. ”*

*We have also added a sentence making this point explicitly in the revised manuscript (lines 432-435): “We acknowledge that the variation in rooting depth across the ensemble is large, but point out that so is uncertainty in plant rooting depth, and moreover that the uncertainty in rooting depth is less well-quantified than other plant traits such as P50, such that this wide variation reflects a real and deep uncertainty in plant rooting profiles.”*

*Moreover, while the importance of deep roots may be obvious to the reviewer, the manuscript goes further than that to show how the sensitivity of other hydraulic traits is mediated by the rooting depth, and thus emphasizes the interplay between aboveground and belowground hydraulic traits using the plant hydraulics model.*

The explanation why the authors chose to change only P50\_gs and not ags is wrong (p11, l297ff; p13, l354ff). P50\_gs tells "only" at what potential close the stomata, eventually. It is actually ags that determines the strategy, i.e. isohydric vs anisohydric behaviour (I think to remember that this is explained in one of the Sperry papers).

*We choose the parameters that are relatively well understood and allow us to catch the safe vs. risky strategies as described by Skelton et al., rather than exhausting all possible parameters in the model, which would come with high computational costs but yields limited further scientific value. In essence, the different combination of p50gs and the shape parameter(ags) can generate similar stomatal response curves. For example, small negative p50gs with small ags would result in flat stoma response curve which is similar to a large negative p50gs combined with large ags. P50gs is well understood and has more observed data, while ags is less studied and has less availability of observed data. Given the formula of FATES-Hydro, with the same shape parameter (ags) value, the slope of the stomatal response curve at P50gs is determined by P50gs in the way that larger negative P50gs results in flatter response curve, and vice versa, which resemble the risky stoma (red curve) and safe stoma (blue curve) in fig 1A in Skelton et al. (2015) respectively. Given that we don't have any measurements of ags at our site, we used the default value of conifers given by Christoffersen et al., (2016). This has been discussed in original MS L340 - 379*

I was wondering at the results section if it would not have been better to adapt the model to the site first. Now it looks like that the model cannot reproduce GPP with any of the hydraulic strategies.

*Yes, we agree that there is a bias in the GPP across these cases. Hydraulic strategy alone certainly will not make the model to match both transpiration and GPP, whereas the latter is more controlled by traits that affect biochemical reaction rate, e.g Vcmax at leaf level and light distribution (or radiation penetration process) at canopy level. We didn't focus the model on the joint calibration of both ET and GPP for several reasons. The first is that we were more interested in exploring the parameter uncertainty and what is said about ecological dynamics of plants with different traits than in calibrating the model per se. Second is that here is some degree of uncertainty in the eddy covariance observations, and that the simulations here are still useful even in the presence of some bias in the GPP predictions. We have explained this in our original MS (L380 - 394 )*



Also, the Psi<sub>l</sub> values are so strange that one should ask oneself if there is not something wrong with the model.

*We assume by “strange”, the reviewer means the leaf water potential is too low which has also been pointed out by the first reviewer. The main reason is the model is run under a static stand mode, the stand density and LAI do not change. This counterfactual experiment is meant to allow comparison of the direct trait effects on ecosystem function, rather than the indirect effects of plant traits on ecosystem composition and function, which is a more complex process occurring over a longer timescale. In reality, the leaves would die when dehydrated. We have discussed the issue in the discussion section (L632 - 636). “Although it might be unrealistic, the leaf water potential can serve as an indicator of the degree of canopy desiccation. With no or very little leaves, trees would rely on the storage carbon to support respiratory demand until the wet season comes to regrow leaves. Depending on the duration of the dry season, trees may exhaust the stored carbon and die from carbon starvation. ”*

Are the curves of Fig. 1 realistic? How do they compare to measurements? Are there PLC (percent loss of conductance) curves for ponderosa pine?

*Unfortunately, we don't have measurement of the vulnerability curve of the pines at that site since they had all died by the time we started this project.*

Why are the K<sub>max</sub> different for the safe and efficient strategies?

*We covaried K<sub>max</sub> with the P50 of xylem because we want to represent the safety-efficiency tradeoff of the xylem system. We have devoted a whole paragraph in introduction section to explain this choice ( L98 – 125)*

*“ The maximum hydraulic conductivity and the vulnerability to cavitation are the two key xylem hydraulic traits. Differences in the anatomy and morphology of the conductive xylem cell structure and anatomy (Hacke et al. 2017) lead to differences in maximum conductivity and the water potential at which cavitation starts to occur (Pockman & Sperry, 2000; Sperry 2003). Within the conifers, there are at least three mechanisms that lead to a tradeoff between xylem safety and efficiency. First is the morphology of the xylem conduit. It is widely acknowledged that narrow (or short) tracheid are safer than wider (or longer) tracheid but have lower conductance per sap area (Choat and Pittermann 2009). Second are the intervessel pit membranes. Thicker and less porous membranes prevent the spread of air but increase the hydraulic resistance of xylem (e.g. Li et al., 2016; Pratt & Jacobsen 2017). The third mechanism comes from the division of limited space (Pratt and Jacobsen 2017). With the same cross sectional area of conduits, vessels with a thicker cell wall provide stronger mechanical support, so that the conduits are less likely to collapse when xylem water potential becomes more negative, however this reduces the area that can be used for conduits transporting water. While these physiological constraints require that the tradeoff does exist to some extent, in many studies, this tradeoff appears to be weak, and there are certainly species that have both safe and efficient xylem. Further, there are many other plant traits that can affect the safety, such as wood density (Pratt and Jacobsen 2017), pit anatomy (Sperry & Hacke 2004, Lens et al. 2011), and biochemistry (Gortan et al. 2011). These traits can have large variations among different plant*

types. The tradeoff will be weakened when grouping plants at a coarse scale, e.g., by biomass, families and/or across a range of geological and climatic region. But when focusing on certain species in a particular region, the tradeoff becomes stronger, as demonstrated by many local studies (e.g. Barnard et al. 2011, Corcuera et al. 2011, Baker et al. 2019). For example, Kilgore et al. (2021) shows that there is a clear safety-efficiency tradeoff across the pine trees in a specific location in the western US. Thus, while we acknowledge that there are many exceptions to the xylem safety-efficiency tradeoff, it is a useful framework for examining plant strategies for dealing with drought.

”

How comes that  $\theta/\theta_{sat}$  increase at the bottom of the soil column in Fig. 5? Is there no drainage (Fig. 6 suggests otherwise) or is  $\theta_{sat}$  very different at the bottom of the soil column?

*This is mainly due to reduced water uptake at the bottom by roots. Because in the FATES the relative proportion of roots decrease with increasing depth. Thus there is less roots to uptake water in the deep soil. The detailed explanation can be found in discussion section L614 - 616 “With deep roots, the vertical soil moisture variation is more homogeneous due to the extensive root distribution. With shallow roots, the soil becomes extremely dry at the surface (<1m) and extremely wet in deep layers (>5m) resulting from the aggregated root distribution in the upper layers.”*

---

I would have lots of further comments but I think the manuscript should be brought into shape first.

*We thank the reviewer for their time and their careful review of the manuscript.*

Just as a last comment on the conclusions: it is very obvious that deeper rooting plants are less influenced by droughts, and that risky and safer xylem is more and less vulnerable to drought but it also has more and less GPP, by definition. These obvious results should not be the main conclusions. And the statement "that deep-rooted pines with risky stomata have the highest GPP but also the highest drought mortality risk" is misleading because shallow-rooted pines with risky stomata actually have the highest drought mortality risk.

*This statement refers to the result that deep rooted trees not refer to all the strategies, only the ones that might survive during the pre-drought period. We will reword this sentence to "that deep-rooted pines with risky stomata have the highest GPP but also the highest drought mortality risk of the plant trait combinations that may survive under non-drought conditions" This point is made in the prior paragraph but for consistency we will make it again here (L760 - 762). “In*

*contrast Whereas, shallow roots with risky stomata leads to high mortality even during non-drought years, thus an uncompetitive realistic combination at that site. "*