

## General comments

This manuscript describes an interesting study that attempts to efficiently parameterize the FATES vegetation demography model in novel ways. The authors start with extant trait observations which are filtered over a number of steps for parameter combinations that produce ecologically realistic forests in which the trait combinations conform to a priori expectations in relation to each other and driving data. This is highly interesting for reasons well described in the text, and I agree with Reviewer 1 about its importance for the broader field and the generally high quality of the presentation and text.

There are a few things that could be improved (see short list below). Specific spots in the text are occasionally awkward or not well integrated; a few of the figures should be tweaked or perhaps re-thought; and parts of the introduction and discussion could be tightened with little loss.

In summary, this is a really interesting approach to a hard problem that should be of wide interest to land modelers generally, and that provides a framework to build on for vegetation demographic models specifically. It needs only minor to moderate revisions before final publication.

## Specific comments

Lines 16-17: not sure this sentence (“Composition is...”) is needed  
We agree and removed the sentence.

53-61: this paragraph feels disconnected from rest of the introduction  
In order to better connect to the previous paragraph, the first sentence now reads  
"Prior research suggests that the model parameters that are most important in determining composition are likely to vary according to the model's representation of the primary constraints on plant growth (Nemani et al., 2003) and disturbance regimes."

Figure 1: this is great—thank you. Extremely helpful in following a moderately complex analysis  
Thank you.

125: “combines observations”  
Corrected

141-142: this (specified rank correlation) is unclear; expand a bit?  
We edited his section to read:  
"We then ordered the sampled parameter values to maintain specified rank correlation between different parameters (Xu and Gertner, 2007; Iman and Conover,

1982). The rank correlation matrix was calculated based on observed trait values for each PFT. Samples for each parameter were drawn from a distribution defined by the observations, such that pairings of samples between parameters maintain the specified rank correlation (Iman and Conover, 1982)."

Figure 2 is really cool. Hard to see light grey versus dark grey though  
The 72-member ensemble points were missing from this figure. They are now plotted and the contrast between light and dark grey is apparent.

Figure 3 caption: "green lines"?  
Changed to "colored lines"

Figure 4: why aren't the canopy and understory outer rings grouped next to each other? I.e. blue canopy -> red canopy -> red understory -> blue understory as one goes around the outer circle  
We use this arrangement because we think that showing the overall composition of pine vs cedar is the most important piece of information in this figure, given that the focus of the manuscript is on forest composition, and that canopy position is secondary.

271: break into two sentences for readability and correct grammar  
Corrected as suggested.

315-, 334-: well described and summarized  
Thank you.

325: kind of circular...perhaps reword  
Edited to: "However, these methods do not ensure that simulated composition, even when accurate, is a result of the mechanisms that determine competitive outcomes and drive composition (Williams et al., 2009)."

345-349: this could be expanded. How onerous \*was\* the computational cost? In the future would multi-site, even complete landscape, workflows be better?  
FATES dramatically increases the cost. We added text noting this, and acknowledging additional site simulations may help: "CLM-FATES simulations are approximately five to six times more computationally expensive than big-leaf CLM simulations. Even so, selecting additional site locations stratified by environmental variables may be beneficial, particularly when developing more than two PFTs."

359-361: a critical point but could be expanded on; is it truly unrealistic, given realistic driving data?  
If the long-term driving datasets of climate and disturbance timing and magnitude are available, we can expect the models to replicate current vegetation distributions. We edited this to read:  
"Current vegetation distributions are the result of particular sequences of climate, disturbances, and dispersal events across millennia (Jackson et al., 2009). Therefore,

without observations of realized disturbances (including land management), and their representation in the model, a global model may not be able to precisely replicate the spatial patterns of vegetation structure and distribution from observations. Functional relationships among climate, disturbances, and vegetation distributions may provide a more meaningful benchmark."