This paper assesses the validity of the carbon allocation scheme in the Community Land Model (CLM4.5) against a range of data sets, primarily from flux towers in North America. Two alternative schemes are also introduced as tested against the same data sources.

In general I found this to be a robust piece of model-data comparison, using novel observations and with high relevance to contemporary issues in land surface model development. It also identifies some significant failings of the existing model and suggests schemes that would lead to improvements. I think the organization of the text could be slightly improved upon, but otherwise I have no major concerns here.

General points:
Throughout the manuscript, there is a tendency to refer to the CLM4.5 as simply 'CLM'. This should be corrected to, wherever possibly, only using CLM4.5, since all other versions of CLM woud produce substantially different findings to those presented here.

In general, much of the writing in this paper leaves the reader in suspense about the purpose of the paragraph until the end, which makes it hard to follow. I suggest trying to reverse the logic of the paragraph/sections, adding the concluding statements much nearer the beginning, and then backing them up with what follows.

Specific Points
P1L34: This 'fixed scheme' is introduced as two seperate schemes, one for evergreen and one for deciduous, but given that (P6L23) these differences are derived only from the model parameters, and thus do not require and structural

modifications, I would see this as one and not two seperate 'schemes'.

P2L10: The last sentence of this paragraph (That coud be done by…) is too vague and should be modified to include specific recommendations.

P2L11-13: This reads as somewhat circular (we need to look at biomass data to get biomass correct)? Maybe remove the very last phrase?

P3L1: Do all LSMs do this? What are the exceptions?

P3L1-17: This long paragraph is somewhat difficult to navigate and could do a better job of clearly introducing the focus of his paper. I suggest splitting it into at least two paragraphs, and making clear what is the focus of this paper early on in each paragraph.

P3L22: What are 'biometric data' in this context?

P4L8: Is there a reference for the L2 product? I wouldn't assume that everyone is 100% familiar with this.

P4L14: Why construct these historical biomass datastreams? Why not just use current biomass estimates? A motivation needs to come before any of this.

P4L24: I'm guessing allometric assumptions have huge impacts on stem carbon estimates and thus on the rest of the results? These should be discussed somewhere.

P5L6: After all the discussion of biomass estimates, more

detail is needed on the leaf C methodologies, since that is non-trivial as well.

P5L9: Not necessary to say PTCLM was used, since this isn't a different model, and makes it seem like it might be.

P5L15: 'dead' stem and root pools are a little confusing. Are these heartwood, or in the litter pool? This needs to be disambiguated.

P5L17: I think all these a1, a2, a3 references should have the numbers in subscript.

P5L30: What is the purpose of fitting these parameters for a broad range of NPP?

P6L24-27: This section seems like results.

P6L30: This sentence make it seem like SLA is prognostic when it is actually fixed.

P7L10: There should be units for these parameters here.

P7L25: What optimization approach? Did you use the adjusted parameters in all of the allocation model simulations, or in a new set of simulations?

P7L27: Is the model emulator 'just' equation 4, or is there more to it that needs explaining?

P7L26: Why are you optimizing turnover? Again, are these new numbers used in all the simulations? There needs to be a better overall narrative connecting the different parts of the

methods section.

P8L5: There needs to be a motivations statement at the beginning of this section, otherwise one has to read all the way through to figure out where the argument is going…

P8L23: These also seem like results.

P9L19: CLM was already introduced much earlier.

P10L6: Why use the CRUNCEP data when these are Ameriflux sites for which meteorological observations are typically available? I don't think this means that the simulations need re-doing, since this paper is focused on relative allocation schemes, but I do think that some more discussion of the potential for errors when comparing site level data and a model driven with reanalysis would be appropriate.

P11L19: This sentence is confusing. State which values - were- used, not which ones were not…

P11L21: I don't understand this 'turnover effect' from this sentence. What are the max and min tem turnover rates?

P12L10: In general this discussion section is rather too long and could do with focusing more coherently on the important findings of the study.

P12L14: What are 'initial' biomass estimates in this context?

P12L16: Couldn't there be different parameters for dec and evergreen plants within the Litton scheme? Why not propose

the ideal scheme within this paper?

P12L28: This sentence is confusing. Why mix the reporting for the evergreen and deciduous forests up like this?

P12L30: Which version of CLM did Hudiberg et al. use?

P13L6: These comparisons with other models are somewhat distracting all the way through this section. I'm not sure it's particularly relevant, given that a) there's no real reason to imagine that there would be a systematic bias and b) the illustrations given are not wide-ranging enough to demonstrate one. I'd suggest moving all that material to its own section or removing it entirely.

P13L5: This seems to suggest that Litton is better but only becasue of the existing biases in NPP?

P13L11: What exactly is being suggested here? I feel like it needs a specific equation.

P14L3: Is the ORCHIDEE bias for the same reason? There are lots of ways to get a high LAI bias!

P14L5L: This root allocation discussion perhaps need to be in its own section (and maybe could be removed since the datasets used here don't really address root allocation per se).

P14L26-: This discussion of 'initial conditions' might also be removed, since 1) this study doesn't really look into initial condition variability 2) it thus doesn't show any sensitivity to initial conditions, 3) what 'initial conditions' are isn't defined

here and 4) the derivation of all of the equilibrium biomass pools earlier terms rather undermines the notion that initial conditions might be important. The IC study used by Kay et al. in particular, illustrates extreme sensitivity to very minor perturbations of atmospheric initialization, but this is not really relevant to the problems presented here.
IC sensitivity is possible in a model like CLM, due to positive feedbacks between low canopy LAI and surface temperature, nitrogen acquisition, etc. but that is not a feature of this analysis.

P15L15: This reference to Xia is confusing. This is with a model other than CLM4.5, but whic one? I'd suggest removing it, since structural modifications of one model are not necessarily relevant to another.

P15L25: What happens at the other sites?

P15L30: This is an important point, maybe highlight more in the abstract, conclusions, etc.

P16L6: This long discussion of Harvard forest rather detracts from why turnover is hard to estimate from the tree ring data? Stem turnover can surely be estimated from permenant sample plot data instead? Further, this whole discussion is really about how plot level observations (rather than the type of observations) are altered by disturbance history. Big leaf models implicitly aggregate all successional stages together, and so comparison with individual sites is problematic, which is a good argument for using site-specific stem mortality estimates. I think this argument could be made clearer, and shorter!

P16L17: The 'large geographical scales' phrase is repeated from several lines earlier .

P16L18: In CLM, land use change is considered separately from natural ecosystem physiology.

P16L21: Published where? And what numbers were used where? This is too vague of a description.

P17L8: I'm not sure that the allocation schemes disagree, they are just different… Can this be rephrased?

P17L6: The dynamic allocation schemes could be interpreted as plausibly operating at a cohort scale, but also could be interpreted entirely at the landscape scale. This secton introduces the idea that cohort representation is needed, but then discusses coherent patterns in the site-level stem/leaf ratios, undermining that argument.

P17L14: Better representation of veg dynamic and functional root representation are very different ambitions for LSMs, and have very different implications for allocation schemes. ED-like models, for example, already use allocation schemes that map onto changing stem/leaf ratios with tree size, but these are inappropriate for big leaf models. Some models (LM3-PPA) already have functional roots and change allocation to the accordingly. It seems like this topic (how to move forward with allocation schemes) is introduced too suddenly in the conclusions, when it might be a better topic for a discussion section evaluating the potential for alternative model improvements to have better connections to data?