This study attempts to improve the mapping of potential natural vegetation in the BLUE model, with the aim of better constraining estimates of emissions from land use change. The study is generally very good; I know about the limitations of DGVMs vegetation succession / distribution schemes, but do not have direct experience with PNV maps – so it was interesting to learn about how this is calculated and the large uncertainties involved. Three main comments and some small ones below.

1) Can the authors show more systematically that their work is an improvement over Hansis? It seems probable; but at present this evidence is scattered across the SI and discussion. I think what we need is a section in the results that draws this together thoroughly. This should contain points where the ML derived map may introduce errors of its own: this is inevitable and should be discussed openly. See discussion below of the Caatinga biome & limits of PFT schemes.
    - E.g. the disaggregation of the "Raingreen shrubs" class seems like a major macro-level improvement. Can you demonstrate that this is so?
    - You also extend the tree line a long way North, removing Abiotic surface cover. This seems to hold in PFT ML Grass. Is there separate / independent evidence for this?

2) Relatedly, subjectivity and uncertainty are unavoidable in this kind of work. However, I think we are missing analysis that assesses how far uncertainties in the underlying data influence the new maps' impact on FLUC estimates. There is a good analysis done with max grass / max tree & Figure 6 is useful – but I don't think this alone is sufficient.

    - What I would like to see is an assessment of how far the model uncertainty data presented in Figure 3 & the underlying scarcity of data used to parameterise the model impact FLUC. E.g., what happens if you only change regions / pixels where you have good data and high model confidence? How much of the increase to FLUC does this remove? I don't know how expensive BLUE is to run, but ideally, we would have some kind of bootstrapped uncertainty range. Otherwise, a single "max confidence" run would be very useful.

3) It would be very helpful for a non-global carbon budget audience to understand what is meant by a "forest" PFT and how this relates to carbon densities. If it is literally being claimed that forest is 100% closed canopy for forest PFTs, then that seems like a big limitation here, particularly in temperate regions. E.g., in Europe and the USA, pre-humans there would clearly have been herbivory, fire

and areas of nutrient limitation that would result in some grass / shrub cover. This is mentioned in passing in the discussion. This may be just my lack of understanding but some description of how the carbon equations in BLUE interpret forest PFTs would be very helpful.

Minor comments

Abstract

"Improves the representation of.." Either be specific about by how much & how you define "improve" or drop this claim from the abstract.

Intro

Lots of good context here.

With model-based assessments, it seems that a major underlying issue in PNV is the equilibrium assumption itself. DGVMs can suffer from vegetation schemes designed around equilibrium assumptions that have later had disturbance tacked on. A few times in the introduction we have general discussion of model issues (e.g. "model structure, complexity, parameterization, input data… line 74) where a little more specificity about the key problematic areas (disturbance being an important one) would be good. Lines 42-44 are another example – what modelling assumptions, what issues in model complexity?

Lines 81-115. It is interesting that you state, probably quite reasonably, that ML approaches reduce subjectivity over expert judgement, but then describe your PNV map as "best guess", presumably due to data limitations (which you acknowledge). Can we reframe this discussion slightly to be clearer about how far we can reasonably remove vs reduce uncertainty/subjectivity?

Methods

Lines 126-127. What years does BIOME 6000 capture? How important is natural climate fluctuation to interpreting the pollen record?

Lines 135-157: I like the fact that you are being clear there was a degree of subjectivity/expert judgement in cleaning the pollen data & assembling them into ML classes. This is what I mean re: lines 81-115. We can reduce, not remove subjectivity.

Lines 172-180: you stratified by biome; but did you account explicitly for spatial autocorrelation in your model validation? Given the extremely skewed spatial sample I think this would help increase robustness. E.g., choosing spatial subsets of your data to hold out for validation rather than random samples by biome?

Line 192; notes the uncertainty is spatially explicit, but I am not clear how? Is biome being used as a synonym for "spatially" here?

Lines 236-244; the sensitivity analysis over PFT distributions/translations is good.

Results

The biome map looks sensible. It does seem that a major limitation is there is much more detail where you have more data. E.g., Europe and the USA have a tapestry of biomes, whilst Africa has big blocks of just "tropical savanna". Can you, I wonder, quantify landscape heterogeneity as a function of data availability?

"Steppe" also occupies very large areas of quite unrelated biomes – xeric environments in central Australia and grasslands in central Canada. Is this a data limitation issue, or is this just that you will allocate these areas to C3/C4 grass and so for your purposes further disaggregation isn't useful?

You describe in lines 365-374 how Southeastern South America moved from tropical deciduous forest to tropical savanna. The Caatinga biome is hard – JULES also struggles with it. But to my mind it isn't a tropical savanna, either. This points to a basic problem of translating biomes into PFTs. How robust is this change in ML max forest, e.g.? Is this regional disagreement between PFT ML and PFT Hansis really about the limits of BLUE's PFT classification?

Discussion

The areas where it is claimed the presented analysis is an improvement (e.g., Patagnoia) should be presented in the results, with figures moved from SI to main text. Same comment 491-499.