Review of Li et al. "Simulating carbon and water fluxes using a coupled process-based terrestrial biosphere model and joint assimilation of leaf area index and surface soil moisture"

Anonymous Referee #2

General Comments

Li et al present a study evaluating the assimilation of LAI, soil moisture, and a joint assimilation of both LAI and SM into the LPJ model. Overall I understand what is happening in this manuscript but in general there is a lot of detail missing that prevents an easy reading of the manuscript and full understanding of the results. The abstract, introduction and methods in particular need improvement – please see detailed suggestions below.

The authors have provided a lot of nice detailed analysis of the three assimilation schemes in the results, which was good to see. All too often in these types of studies there is only a rudimentary evaluation of the assimilation that is performed. I commend the authors on this. However, I would also suggest that at the beginning of each section or even each paragraph they include a summary overview sentence to explain the key points of that paragraph/section, so the reader knows what to expect and look for when digging into the detail.

I liked the key results and discussion relating to the difference between humid and arid areas (with the exception of one comment below). These points could be highlighted even more by the authors.

Detailed suggestions

Abstract

Confusing mix of LPJ-*something* used here. First LPJ-Vegetation, then LPJ-DGVM, then LPJ-PM. Also LPJ-VSJA, but I appreciate that is the DA system (although the VSJA acronym is not explained). Then in line the introduction the authors talk about LSMs, not DGVMs and at line 66 terrestrial biosphere models are mentioned. Please be clear and consistent throughout the manuscript.

Please explain all acronyms. Once you've explained an acronym then use that throughout.

A clear explanation of which is the model version that has been optimized with the DA framework and which not would be helpful in the abstract.

Line 34: "The assimilated GPP and ET" suggests that GPP and ET data have been assimilated. I suggest "posterior GPP and ET" would be better.

Introduction

Line 65: Probably more appropriate references here. See Scholze et al. (2017) or Exbrayat et al. (2019) for further references.

Line 68: You also need the underlying model, not just these three components.

Line 71-73: I would re-write this sentence as "which significantly improve simulations by periodically updating state variables (e.g., LAI and soil moisture) using remote sensing data without changing the model structure".

Line 74: "obtain the dynamic balance of the estimation window" ï□ I would explain fully what is meant by this for non DA specialists. It might also be useful to add an additional sentence explaining the difference between EnKF and 4DVar either before or after this set of sentences.

Line 79: Please can the authors be more specific when they say "satisfactory performance in land DA" beyond what is specified for a different paper later in the sentence (that the method does better at estimating GPP and ET with ENKF)?

Line 85: I am not sure you want to reference Liu et al here because they talk about how different LAI products have inconsistent estimates; therefore, that is a disadvantage for using LAI data to evaluate or optimize models, as how do we know which LAI product is more accurate? This actually is in contrast to lines 94-96.

Line 88: Do the authors mean more accurate SM data *assimilated into models* can improve accuracy? And if the authors are not talking about assimilating SM data here, then how was SM data used to improve accuracy of models and is that relevant to a DA study? Same comment for the references used on Line 85. From the sentence they're referencing I assume these references demonstrate how LAI has been used to improve models, but I am not sure that is the case. If instead these references are to demonstrate uncertainty in these variables in mdoels then that should be better specified.

Line 104: Maybe the authors could explain why microwave RS instruments are used to detec soil moisture, and how that differs to the type of RS instruments that are used to derive LAI data, for the purposes of consistency.

Lines 122-124: Do the authors imply that they are assimilating global data, i.e. every grid cell of the products? This needs to be made clearer in Section 3.2. There have been other studies assimilating LAI and SM, even if they have not See Wu et al. (2018) as well as other papers from the same authors/group as the Bonan et al. (2020) paper. The introduction needs to be expanded beyond to reflect this history and how this study builds on that beyond just the assimilation of global data. Or at least, their hypothesis for how the assimilation of global data will be a step beyond those previous studies, but that that

hypothesis needs to be evaluated in their analysis/results. In short, the authors need to do better at explaining, or demonstrating via analysis, why their study goes beyond the previous land DA studies assimilating LAI and SM. The authors need to answer the question "what do we learn from this study beyond what past studies have told us?". Points could be added to discussion too. This will help the modeling and DA community more widely discern the best practices and possible pitfalls for assimilation of these two datasets. If it is purely a technical advance (e.g. sheer scale of obs etc), then those advances and lessons learned should be highlighted more in this manuscript. The authors could add specific questions that they are trying to answer to the final paragraph of the introduction.

Methods

Table 1: Is LPJ-VSJA used for assimilating data into LPJ-DGVM or LPJ-PM? I would have thought the latter?

Lines: 147-149: Not sure I understand here. There is or is not soil stratification in LPJ? And please could the authors explain how that connects to simulating water limited regions? I also think this sentence might be better after the authors have explained LPJ more generally.

Line 152: Need much more information than this: "the GPP is calculated by implementing coupled photosynthesis and water balance" with references.

Lines 147-161: I feel like the reader needs a lot more basic information on LPJ and the PT-JPL models. Perhaps they could have their own sections before describing how, and why, the models are combined?

Line 167: What do the authors mean when they say "The SMAP SM was applied to model global ET using PT-JPLSM"? Do they mean the data was assimilated?

Line 170: The authors talk about "scheme 2" here before talking about scheme 1? This is confusing. Please resolve.

Line 169-176: I am a bit confused by what is going on in this paragraph. Please make it more clear for the reader.

Line 185: Earlier you say "PODEN4DVAR".

Lines 190-205: This whole paragraph is difficult to parse as there are no sentences and instead there are a confusing number of semi-colons. I know the authors are describing steps, but I strongly encourage them to split this up into sentences. You can always start a sentence with Step 1 or Step 2 etc.

Lines 201-202: which dataset did the authors use to define humid, semi-arid etc?

Line 210: "propagated by energy transmission and ecosystem processes in the dynamic model

" ï□ Not clear what this means.

Lines 216-217: How were these model and observation error % chosen? The authors should provide an explanation and references. Same for lines 234-235.

Lines 218-221: What are the scale factors? What are the integration members? These have not been explained. I am confused again at lines 246-249.

Section 2.2.3: this is really a step-wise assimilation, rather than a true "simultaneous" joint assimilation. There are are advantages and disadvantages to that that should be discussed, and assumptions explained. See MacBean et al. (2016) for discussion.

Line 244-245: "Finally, GPPCO and ETCO were output by joint assimilation based on the POD-En4DVar method." ï□ I am confused here. This sentence reads like a separate joint assimilation is done when from earlier in the section/paragraph it seems like the LAI and SM/ET have already been assimilated?

Line 251: Earlier you said the "PODEn4DVAR" reference was Tian and Feng 2015.

Line 252: Explain what "POD base" is. And at line 269 please explain "POD decomposition".

Line 254: "flow-dependent error estimates" ï□ please explain what this is for the non DA specialist.

In general the number of subtext acronyms is difficult to parse. I suggest the authors find a slightly different way to refer to all the variables. For example GPP_prior, GPP_scheme1, GPP_scheme2 etc.

Results

Figure 2: Hard to tell what a,b,c,d is for each set of metrics just by looking at the figures. Figure text should be larger.

Figures 3 and 6 are not referenced in the text.

Figure 8: would be useful to put the labels "semi-arid" etc inside the actual subplots.

Lines 468-476: this is nice but it would be great to see the prior model-data comparison to see how the "CO" optimization has improved things. Otherwise, the authors' claim at line 476 that SM data are needed for water-limited areas is an overreach. Actually, without comparing to schemes 1 and 2 it is hard to say whether it is SM or LAI data that have achieved a good result in water-limited areas. The authors do seem to discuss the prior in the paragraph lines 485-490 but I am having trouble seeing where this fits into the bigger picture.

Line 496: do you mean Figure 7 here?

Figure 9: GPP_SM and ET_PM? I am confused here? Labels on the subplots would also help here.

Figure 10: the color coded grid is helpful here.

Discussion

Generally a well-rounded discussion of the advantages and caveats of the approach. I would appreciate more discussion on the inconsistency between LAI products in Section 5.3, and implications of the fact the assimilated products (LAI and SM) may be biased. What impact do the authors think that would have on the results? Also issues related to temporal sampling interval could be discussed somewhere in the discussion, as well as assumptions/caveats of the DA method that may affect the results.

References

Bonan, B., Albergel, C., Zheng, Y., Barbu, A. L., Fairbairn, D., Munier, S., & Calvet, J. C. (2020). An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: application over the Euro-Mediterranean region. *Hydrology and Earth System Sciences*, *24*(1), 325-347.

Exbrayat, J. F., Bloom, A. A., Carvalhais, N., Fischer, R., Huth, A., MacBean, N., & Williams, M. (2019). Understanding the land carbon cycle with space data: current status and prospects. *Surveys in Geophysics*, *40*(4), 735-755.

Scholze, M., Buchwitz, M., Dorigo, W., Guanter, L., & Quegan, S. (2017). Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems. *Biogeosciences*, *14*(14), 3401-3429.

Wu, M., Scholze, M., Voßbeck, M., & Kaminski, T. (2019, January). Combining SMOS soil moisture and JRC-TIP FAPAR for better constraining global carbon fluxes during 2010-2015 within CCDAS. In *Geophysical Research Abstracts*(Vol. 21).