Comment on bg-2021-248

Anonymous Referee #1

Referee comment on "Assimilation of passive microwave vegetation optical depth in LDAS-Monde: a case study over the continental USA" by Anthony Mucia et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-248-RC1, 2021

Assimilation of passive microwave vegetation optical depth in LDAS-Monde: a case study over the continental US

Anthony Mucia et al., 2021

The paper illustrates how atellite-based vegetation data assimilation, and joint vegetation and soil moisture assimilation has an impact on evapotranspiration, gross primary productivity and soil moisture over CONUS. The paper is of general interest to the scientific audience, but needs some clarifications and could also benefit from a careful language check.

Major:

1. The intense massaging of the VOD data begs the question which signal is ultimately helping the assimilation system in this paper. It would be good to make this clearer in the manuscript.

1a. The rescaling of VOD to LAI needs to be spelled out more precisely. Is a relationship found pixel per pixel and per season, or was the relationship based on the CONUS clouds per vegetation class for all months or only for the growing season or per season presented in Fig 5, or still something else?

Unless I missed it, there is also no mentioning of how the discrepancy in spatial resolution (and the spatial-temporal collocation in general) in VOD and LAI data is handled to obtain this linear relationship.

1b. After rescaling and applying a 90-day rolling average, the short-term variability is probably gone. Yet, the latter might be very important to catch the start of the growing season. In the end, it sounds like the only 'original' signal that can improve the assimilation system is of an interannual (and perhaps seasonal) nature. If that is correct, it needs to be explicitly mentioned in the paper. Finally, the 90-day rolling average means that the benefit of a filter is defeated: why not simply directly use a smoother and limit the observation preprocessing?

2. Related to the above, the choice to evaluate the results only in terms of Pearson correlation needs to be explained. Is it not more common to evaluate at least soil moisture in terms of anomaly correlations? And how about including an evaluation in terms of unbiased RMSD, or at least mention if the story remains the same for other metrics?

3. The monthly correlations (fig 7, 12) are not clear: (i) are all time steps included, i.e. both forecasts and analysis time steps, and at which temporal resolution, (ii) are these values spatial correlations between simulations and reference data, temporally averaged per month, or (iij) are these values multi-year temporal correlations computed at each location and then spatially averaged?

It would be nice to also

- compute confidence intervals for these monthly correlations;

- show the number of pixels involved per month (the high correlations for SSM in the winter month might be applicable to far less pixels, if QC screening was applied)

Minor:

1. Please read the manuscript thoroughly another time. There is some imprecise language and there are plenty of grammar issues. Random examples are listed here (the paper is full of issues; far too many to start noting):

L. 5-L.8: This capability -> this positive impact (implicit flow of thoughts)

L. 5 difference between model simulations and forecasts (drop forecasts?)

L. 8: due to the low temporal…, [which is] at best [,] every ten days, and can suffer

L.13: far more .. than.. product*s* (or *an* optical product…)

L.110: for nature tiles. What is "nature"? There is a hint on line 135, that the model converts urban to bare rock – and rocks are nature?

L. 119: NIT option not explained.

L. 176: LAI that has been of direct estimations… (rephrase?)

L. 182: large -> long wavelengths

L. 192: VODX and VODC are not known to everyone, introduce

L. 290: mention spatial resolution of simulations?

L. 387: , and all the model, and even all the observations… (rephrase)

L. 553: first sentence is poorly constructed, rephrase.

Throughout: use the same number of significant numbers in the text and figures (we have everything from 0.8, to 0.66 to 0.795 for R-values)

2. L. 58: Assimilation here assumes a dynamic vegetation model, which is not present in all LSMs. In the broad sense, LAI assimilation could also refer to an updating of input LAI parameters.

3. L. 107: is the same 20% error applied to actual LAI observations and VOD observations that are rescaled to LAI? Or did you 'rescale' the observation error somehow? Figure 4 implicitly shows that the observation error (relative to the model LAI) will be different for both. It would be nice to check the error between the model LAI and the observed LAI and the LAI-rescaled VOD and at least correct the observation error accordingly to interpret the results.

4. L. 156: hard to believe that the CCI product provides *daily* data from 1978 onwards. If so, then some interpolation must have happened, and it would not be recommended to assimilate interpolated data.

5. ALEXI and FLUXCOM both use MODIS LAI-related data at some point. Would you expect even more consistency with these 'reference products' when assimilating MODIS LAI? What is the issue about data access for FLUXCOM? (this is really for the FLUXCOM developers - I want to raise awareness for open data access)

6. The text jumps from Fig. 9 to Fig 12; Fig 10-11 are only discussed later. Re-order the figures; perhaps the latter figures can even be removed and be presented in a table (~ table 4).

7. The impact on SSM is negligible in this paper and not all in line with other studies. Is the system designed to minimally update SM, i.e. to avoid harm? How general is the conclusion that vegetation DA has a greater impact? Is it just for the ISBA model or would you expect it to be general for all LSMs?

8. L. 505: why is there a discussion about L-band VOD if no L-band VOD is used in this paper? Similarly, why is section 4.3 in this paper?