

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-236-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-236

Anonymous Referee #2

Referee comment on "Development of an open-source regional data assimilation system in PEcAn v. 1.7.2: application to carbon cycle reanalysis across the contiguous US using SIPNET" by Hamze Dokoohaki et al., Geosci. Model Dev. Discuss.,
<https://doi.org/10.5194/gmd-2021-236-RC2>, 2021

This paper describes the author's attempts to parameterize a simple carbon cycle model as two PFTs at 6 flux tower sites, and run the parameterized model at 500 locations across CONUS for 33 years, assimilating estimates of above ground biomass and leaf area on 15 July annually. This represents a complex, technological feat, and the successful development of impressive workflow capabilities. The authors acknowledge that this work is "proof of concept" and that there is a companion paper in preparation. Perhaps this paper suffers as a consequence, as in its current form it offers only very limited progression from previous work.

Currently, the results as presented do not really represent a "reanalysis across the contiguous US". Indeed, there is no discussion of regional/continental scale biomass or flux estimates at all, which is surprising given the article title and abstract. It does seem like these should be amended to reflect the end point of this current study more accurately.

The authors suggest that it is technically very challenging to produce a detailed land surface reanalysis using a complex carbon cycle model contiguously at regional or continental scales. Nonetheless, this is now routinely done by multiple different groups (e.g. Albergel et al., 2017, 2010; Bacour et al., 2015; Boussetta et al., 2015; Demarty et al., 2007; Kumar et al., 2019; Ling et al., 2019; Raczka et al., 2021) and this work needs to be acknowledged, and critiqued if applicable.

Given that it will be a companion paper in preparation that will describe a regional/continental scale analysis, the question then becomes what is the contribution of this paper beyond the excellent works of Fer et al 2018 and Raiho et al 2020, which thoroughly describe the methodologies employed here?

This could be a detailed technical assessment of skill, or otherwise, of the model and data assimilation system. The results and discussion does begin to do this, but should contain a range of additional, rigorous metrics that are routinely used describing the performance of both parameter estimation and data assimilation – for example prior ranges and posterior distributions on parameters, the number of observations available for assimilation, the number rejected due to QC, measures of bias, histograms of residuals, etc. These, and others, are required to be able to judge the performance of the system and move beyond qualitative description. For example, suggesting that “SIPNET showed considerable improvement” (line 341) after optimization seems quite meaningless, not least because at some sites model performance doesn’t improve at all. What is the cause of the marked variability in improvement between sites?

The spread in Figure 6 and the remaining biases after optimization in Table 1 seem to actually call into question whether this model is able to produce meaningful results *at all* when trying to represent a “PFT”, rather than a single site location for which it was originally designed. This needs to be thoroughly discussed.

Suggest that “Freerun” or “Open-loop” is used rather than “Forecast” in Figures 7 and 8 to make it clear that this has not been constrained by the DA system from an initial condition.

The interpretation of Figures 7 and 10 that the addition of additional data streams has little impact seems surprising. The authors suggest that the “Analysis LAI contracted considerably around the observations”. In fact, the LAI analysis expands to accurately represent the prescribed observation error, as it should in a well-functioning system, but the ensemble mean approximately halves. A general interpretation of this would be that the model partitioning between leaf and other ABG carbon is incorrect. It is really impossible to draw any conclusion from Figure 10 without further diagnostics describing ensemble mean values, and error and bias across the 500 sites.

The additional of more detailed diagnostics would also help to describe and assess how the system performs in modeling fluxes across CONUS. Currently, there is no discussion of where and when it does well or poorly. The authors indicate that 46 locations are selected to allow validation of fluxes (line 204), but no results of this are presented, and this is absolutely required.

The “Future Directions” section could be a lot more focused. It presents a long list of general possibilities that could be applicable to any terrestrial carbon cycle DA exercise – better models, more observations, higher resolution etc. Which, if any, are the authors actively pursuing with this system, and what is specific about their system that makes those developments particularly relevant?

The appendix describing localization is interesting, and represents an additional development in methodology beyond that already described by Fer et al. and Raiho et al.

Typically, it might be only expected to play an important role in model locations without observations – which by design is not the case here (assuming not many observations fail QC, which is needs to described and discussed as part of wider additional diagnostics as suggested above) but it's actually role here should be described, along with the implications of doing the localization by straightforward distance, when in actuality the model domain is being discretized along multiple environmental axes by the authors sampling technique. What is the implication of doing that for your localization approach?

References:

Albergel, C., Calvet, J.-C., Mahfouf, J.-F., Rüdiger, C., Barbu, A.L., Lafont, S., Roujean, J.-L., Walker, J.P., Crapeau, M., Wigneron, J.-P., 2010. Monitoring of water and carbon fluxes using a land data assimilation system: a case study for southwestern France. *Hydrol. Earth Syst. Sci.* 14, 1109–1124.

Albergel, C., Munier, S., Leroux, D.J., Dewaele, H., Fairbairn, D., Barbu, A.L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Moigne, P.L., Decharme, B., Mahfouf, J.-F., Calvet, J.-C., 2017. Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area. *Geoscientific Model Development* 10, 3889–3912.

Bacour, C., Peylin, P., MacBean, N., Rayner, P.J., Delage, F., Chevallier, F., Weiss, M., Demarty, J., Santaren, D., Baret, F., Berveiller, D., Dufrêne, E., Prunet, P., 2015. Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model: JOINT ASSIMILATION OF NEE, LE, AND FAPAR. *J. Geophys. Res. Biogeosci.* 120, 1839–1857.

Boussetta, S., Balsamo, G., Dutra, E., Beljaars, A., Albergel, C., 2015. Assimilation of surface albedo and vegetation states from satellite observations and their impact on numerical weather prediction. *Remote Sens. Environ.* 163, 111–126.

Demarty, J., Chevallier, F., Friend, A.D., Viovy, N., Piao, S., Ciais, P., 2007. Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model. *Geophys. Res. Lett.* 34, L15402.

Kumar, S.V., M. Mocko, D., Wang, S., Peters-Lidard, C.D., Borak, J., 2019. Assimilation of Remotely Sensed Leaf Area Index into the Noah-MP Land Surface Model: Impacts on Water and Carbon Fluxes and States over the Continental United States. *J.*

Hydrometeorol. 20, 1359–1377.

Ling, X.L., Fu, C.B., Guo, W.D., Yang, Z. 2019. Assimilation of Remotely Sensed LAI Into CLM4CN Using DART. *J. Adv. Model. Earth Syst.* 11, 2768–2786.

Raczka, B., Hoar, T.J., Duarte, H.F., Fox, A.M., Anderson, J.L., Bowling, D.R., Lin, J.C., 2021. Improving CLM5.0 biomass and carbon exchange across the western US using a data assimilation system. *J. Adv. Model. Earth Syst.*