

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-190-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-190

Anonymous Referee #2

Referee comment on "CARDAMOM-FluxVal version 1.0: a FLUXNET-based validation system for CARDAMOM carbon and water flux estimates" by Yan Yang et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-190-RC2>, 2021

This study, "CARDAMOM-FluxVal Version 1.0: a FLUXNET-based Validation System for CARDAMOM Carbon and Water Flux Estimates" by Yang et al. provides benchmarking for the performances of the CARDAMOM-DALEC modeling framework across FLUXNET (2015 release) sites. The manuscript focuses on the performances of GPP, NEE and, ET fluxes with only RS-based constraints (A2) and with additional FLUXNET observations of the same variables (A1). The manuscript shows that the performance of the model improves, in general, across all variables and temporal scales. The added CARDAMOM-FluxVal workflow, therefore, definitely brings in a useful additional data stream to the CARDAMOM-DALEC framework, and may thus be a valuable baseline for scientific studies of a similar nature.

Nevertheless, the study and manuscript would benefit by being a bit more comprehensive (with additional analyses or detailed discussions) about the following aspects.

- As mentioned in the limitations, the use of nearest gridded (a half degree) ERA-Interim (hereafter, ERA-I) data for forcing the DALEC model is a major limitation. But, the rationale and necessity of using the gridded data instead of potentially using the tower measurements are not clearly described. It may be that it was not possible because there are gaps in eddy tower-based meteorological measurement, but this has to be mentioned and discussed. Also, one can provide a meta-analysis on comparison of (the available) meteorological variables from the tower and the corresponding ERA-I estimates for the sites. Such analysis would provide information on whether the gridded data are representative of the ecosystem micro-climate or whether they can be used at all. Additionally, it seems that Pastorello et al (2020) also provide the gap-filled meteorological variables. Was this option evaluated as well?
- The baseline remote sensing constraints: The paper mentions that the MODIS LAI was aggregated from the original 1 km resolution. It was unclear if it was aggregated to a

half degree or some other resolution. For the site level simulation, the 1 km resolution data would probably be the closest to the normal footprint of eddy towers. So, I suggest discussing why such aggregation would be needed. Perhaps, aggregating the LAI to a coarser resolution may make it more consistent with the ERA-Interim climate at the same resolution. This may explain why the baseline A2 simulations are already performing quite well. One specific question would be, "does the half-degree forcing reproduce the LAI variability at 1 km resolution or does the LAI have to be aggregated as well?"

- Use of the gap-filled observation: It was unclear why the gap-filled variables from FLUXNET were used. Any particular reasons to use gap-filled constraints were not given, and the cost metric (likelihood) can be calculated only using the time steps which have the observations. The manuscript would benefit by having an explanation of how/why we pick the right observational data variable when optimizing model parameters. In my opinion, the gap-filling itself can be a source of uncertainty. For example, is there a systematic pattern between the fraction of gap-filled observation and the (lack of) improvement in model performance?
- Consideration of observational uncertainty: In the likelihood function, each data stream has an error scaling (σ , sigma). It is mentioned that it represents error across model and data. But it is unclear how these values were set. Was it based on the observational uncertainty (available for some of the FLUXNET variables)? How does σ (sigma) affect the parameter inversion or a contribution of a particular data stream to the total likelihood?
- PFT-level comparison: The manuscript presents a PFT-level analysis of the model performances. As the parameters are optimized per site (I assume), it is unclear if the performance should be associated with PFT at all. Are there PFT-specific parameters or are the relatively poorer performance in non-forest PFT indicative of model structural shortcomings? Also, would climate-based segregation reveal anything interesting? ET would be a much valuable constraint in a moisture-limited climate than in an energy-limited one.
- Parameter uncertainty: The manuscript presents an analysis on parameter uncertainty and model performances across sites, but falls short in addressing a more basic question of whether including additional FLUXNET data-stream helps reduce the uncertainty of estimated parameters within a site. For example, the parameters related to (radiation and) water may be better constrained when an additional data stream of ET is introduced. The manuscript can be better in this aspect. Table S4 should be extended to include the optimized parameter values and uncertainty ranges for both A1 and A2 experiments. As the study clearly states in the introduction and motivation, additional data streams can not only help in improving the performance of the model but also potentially reduce parameter uncertainty and identify model structure errors. The first aspect is well covered, but discussion on the last two aspects would be equally useful as well.