

Review “A comprehensive land surface vegetation model for multi-stream data assimilation, D&B v1.0”

General remarks

The authors have responded well to the reviewers’ comments and improved the quality of the manuscript. However, some issues requiring clarification remain.

Major comments

1. Introduction: I have some trouble understanding what kind of data the authors want to use for data assimilation in the future. After rereading the motivation several times it seems to me that they aim at incorporating remote sensing or earth observation data for variables that have not or barely been processed (e.g. FAPAR) but not for variables that are derived by a model (e.g. GPP) because the second are not well constrained (L30f). If this is the case this is at least to me not easy to comprehend. If this is not the case and the authors plan on also using heavily processed data such as GPP, how will using such badly constrained data improve the process-based model? Can you please clarify?
2. L496-499: Why do the authors think SIF is a useful metric for model evaluation if at the same time they state that the observations show a high uncertainty and they therefore have to increase their scaling parameter s_{SIF} – which mainly seems to be a tuning parameter (see L602f) - by an entire order of magnitude? Considering the large uncertainty I would imagine that using a Bayesian approach, the posterior parameter distribution of a model after data assimilation is not well defined. Are there alternatives to SIF observations that have a lower uncertainty?

Minor comments

1. Check correct usage of “than” and “then”.
2. L74: I suggest to add one or two example variables that directly link to remote sensing information to make the connection to the previous paragraph.
3. L218f: “[...] wood directly feeds SOM.” What is the reason for this and why is this an acceptable simplification?
4. L227: What is meant by “the footprint of the observations”?
5. Eq. 11: What are \bar{f}_{mod} and \bar{f}_{obs} ? I would assume that similar to Eq. 9 the authors would use f_{mod} and f_{obs} and not the annual average.
6. L375: “form” should be “from”.
7. L409-412: The sentence in brackets is very long, hard to comprehend and potentially wrong. It states “[...] model-measurement correlation for NEE to be zero, but substantial [...]”, but how can something be zero and substantial at the same time. I suggest to split this into at least two sentences.
8. Sec 5.1: This section only refers to the boreal site, which could be stated again at the beginning of the section.
9. Sec 5.3.: This again does not include a statement regarding the reported site.