Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC3
https://doi.org/10.5194/bg-2022-160-RC3, 2022

Comment on bg-2022-160

Anonymous Referee #3

Referee comment on "Scale variance in the carbon dynamics of fragmented, mixed-use landscapes estimated using model–data fusion" by David T. Milodowski et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2022-160-RC3, 2022

Dear authors,

I feel like this study and methodology of cardamom represents a major advancement in model calibration. It is particularly exciting to see a framework that could run autonomously using earth observation data. The reproducible nature of this data fusion and calibration process, when coupled with the Bayesian methodology for error estimation (and propagation) could provide more iterable forecasts of carbon or other ecological processes. The question of how spatial heterogeneity plays out in cellular automaton models with single calibrations is crucial to future forecasting and I appreciate the focus on both natural and anthropogenic disturbances. Further, I appreciate the authors' efforts to address the problem of Jenkin's inequity and the role scale selection plays in informing the forecasting and responding. I think the authors' use of stratification by land cover represents a relatively straightforward, logical, and widely available method by which to create more representative models.

However, I feel that some of the conclusions may overreach their results (particularly without independent evaluation). While this methodology has the possibility to improve the estimation of fluxes (etc.), it is not validated to have done so against measurement. The paper presents a gap in understanding to what degree model performance was improved while making some strong claims of the level of ecological fidelity it is able to preserve. I feel this study is both novel and relevant. I would like to see the authors either change the language and address more of the existing limitations of this study or provide further validation to some of the authors' larger claims. I look forward to receiving your response and want to thank you for conducting great work.

General comments

The problem of Jenkin's inequity in landscape or earth systems modeling is a valid and underrepresented viewpoint. However, the answer the authors' model provides can not solve this. Raising this in the introduction raises the idea that authors' methods will be solving or improving on the current structure. Please address in the discussion whether authors feel results provide further proof for Jenkin's inequity or whether they work to address it. This paper does not clearly quantify the advantage of picking one scale (sub-degree +LUC) in a scale variant system. Some may seem likely or self-evident but would

need to be proven. For example, the effect of Jenkin's inequity might be quite similar from the cellular to sub-degree+LUC scale when compared to the explicit plant scale. The authors do however do a great job quantifying that there is an amount of scale variance in this model. Quantification and discussion of how this scale variance impacts forecast, when compared to observable phenomena, would provide a lot of support to this paper.

If I understand the authors' methods correctly, for the baseline model the authors parametrize each pixel separately and in the stratified version the authors separately parameterize each pixel and each pixel's land cover. Is there no information shared across pixels or land cover? Given that land covers would presumably share ecological properties, what is the advantage of not using a hierarchical Bayesian, with priors informed by the larger population of land covers or a bayesian mixed model approach? Please either clarify the decision to make this choice or discuss further the limitations of the separate pixel approach.

Given that none of your parameters seem to directly map onto disturbance, is your model capturing the heterogeneity on the landscape, or overfitting a model? Perhaps some of my concern comes from a lack of understanding of how harvest or fire operates in this model (see below). Harvest would be inversely correlated with the likelihood of future harvest at certain temporal scales and correlated at others. At the scale of a few decades recovering stands would likely (though not exclusively) experience an increase in GPP as forests regrow. Given the stationarity of your parameterization, how does your parameterization constrain such instances?

Hypothesis three- could be improved. To test that any two methods of model parameterization will have contrasting parameters is almost by definition true. Further, they will always have divergent projections on some level. Please provide better constraints to make this hypothesis falsifiable or use a more stringent definition of the contrasting and divergening of parameters.

Specific comments

Line 155: Please provide either here, in the results, or in the appendix the results of the MCMC process. Or how your criteria for model convergence. Accepted sample rates, plots of autocorrelation, and hyperparameters provided are all necessary to determine if confidence intervals are reasonable.

Line 167: Given the importance of EDCs in determining this you should list them plainly, and discuss the constraint they do or do not provide with regard to the function you are trying to achieve.
See: Buotte, P. C., Koven, C. D., Xu, C., Shuman, J. K., Goulden, M. L., Levis, S., ... & Kueppers, L. M. (2021). Capturing functional strategies and compositional dynamics in vegetation demographic models. *Biogeosciences*, *18*(14), 4473-4490.

Line 256 (Disturbance): My apologies if I misunderstand this in other comments. Can the authors please provide greater detail on how disturbance is implemented in the model? This paragraph deals primarily with how it is constrained. There are no direct parameters listed in table A1. If I interpret this correctly, did the authors remove a percentage of tree cover or carbon % to match these data sets? Again, given that this is one of the key differences in the stratified model, a better understanding of the disturbances function in the model is important.

Line 320: Shredding of information implies a specificity not realized here. Information loss is inherent in all models. Without estimating the level of information that is lost, shredding seems overly evocative.

Line 361: While more consistent, what evidence do we have that the prediction is significantly or functionally different from the baseline model? The confidence intervals seem to overlap significantly.

Line 376: I feel this sentence speaks to my larger concerns. There is no way to say that disaggregation ensures the ecological fidelity of a system. Ecology is also scale-dependent. Further, without validation by observation, there is no way to know that the version outperforms the previous version, given that Jenkins inequity would be a property of this scale as well.
Line 385: Different modeling frameworks providing different (though I would not say divergent) outcomes are highly likely. I feel your argument would be improved if you would better quantify or qualify the significance (either statistical or practical) of this level of difference.
Line 410: Given static and statistical parameterization, it would be nice to understand the climate change implications of the stratification approach.

Line 436: The more you stratify a single cell, the greater proportion of it would be captured by this edge or gradient space. If the gradient space has unique ecosystem properties, is there a point where further stratification would further miscalibrate the model?

If helpful, see: Cushman, S. A., Gutzweiler, K., Evans, J. S., & McGarigal, K. (2010). The gradient paradigm: a conceptual and analytical framework for landscape ecology. In *Spatial complexity, informatics, and wildlife conservation* (pp. 83-108). Springer, Tokyo.

Line 459: Again, accounting for subcellular processes at the scale you provide by stratification likely also has a high amount of ecological information loss.

Line 467: While conceptually likely that this provides improved flux estimates, I don't think you have provided enough validation to show this is true. Reduced parameter

uncertainty does not dictate estimation capability. Also, if I understand section 3.1 correctly then the parameter uncertainty is roughly similar, though the means may converge indicating some level of reduced scale variance. That this method reduces scale variance, does not directly imply improved estimation.

Figure 9: I feel that this figure is crucial to your larger argument of scale-dependent outcomes impacting future projections. I feel several aspects of this figure should be revised. Do these model runs represent the median trait estimation or a single draw of the cardamom traits? Please explain in the text. Further, why is the error not propagated here, given the Bayesian approach? This seems crucial to the case that these methods result in fundamentally different models. It is hard for me to understand the implementation (or lack thereof) of disturbance in these forecasts, given that none of the parameters presented would represent that explicitly. See the above comment, some of this may be a misunderstanding of how disturbance works within the model. If the disturbance is only applied top-down, do these projections represent what you captured (that disturbance is highly scale-relevant)?