Response to Reviewer #3:

D. T. Milodowski, T. L. Smallman, and M. Williams

I thank the authors for their revision's, I believe the language is now more appropriate to the findings, with appropriate caveats. I appreciate the additional figures in the supplemental.

We are happy to hear that the reviewer is satisfied with the changes made during the last revision.

Line 20: Please clarify it reads "The differences in simulated disturbance fluxes of terrestrial carbon balance that suggest a C sink in the stratified experiment is weaker than in the stratified experiment." Should one of these be the baseline experiment?

We thank the reviewer for highlighting this error in the text. We have corrected it so that the sentence now makes sense.

> *"The differences in the simulated disturbance fluxes result in estimates of the terrestrial carbon balance in the stratified experiment that suggest a weaker C sink compared to the baseline experiment."*

Line 125: Per the comment in my previous review. I still don't know how you test the hypothesis that the parameters will be degraded, without a validation or a known true parameter value. In response to some of my prior comments, the logic is circular; that aggregating degrades the ecological value and so we will test if aggregating the data degrades the ecological value of the parameters. Could you use "there will be greater error in parameter estimations when assimilating data streams at coarser resolutions" or that component sections will distinctly differ from their aggregated means (as is the case in this study with the minority land types)?

We thank the reviewer for highlighting that this hypothesis statement can be improved.

We have paired experimental tests across spatial resolutions, one in which we simply average EO signals across larger pixels, and another in which we stratify based on functional variations in land cover at fine resolutions before aggregating. In both cases, there is inevitable information loss associated with aggregation, however, what is clear from the pairwise comparison is that the baseline simulations demonstrate a clear contraction of parameter estimates towards some "average" parameter value. In contrast, the distributions of parameters retrieved in the stratified analysis are much more consistent across spatial resolutions. To the extent that the parameters carry information about ecosystem function (notwithstanding the important potential issues around model structure, imperfect data etc., which provide additional sources of error), we can say that this represents a loss of functional information in the retrieved parameters. The only difference in the two approaches is the way in which the data are aggregated prior to data assimilation. Thus, the systematic shifts observed in the baseline experiments are predominately driven by averaging the signal to across different land cover types as spatial scale increases. The issue then is not simply that the retrieved parameter estimates are erroneous, but rather that a single parameter set cannot provide an appropriate representation of the functional variation. This is an important consideration when interpreting the parameter retrievals of large-scale spatially explicit data assimilation studies, which necessarily have coarse spatial resolution relative to the functional fabric of the underlying landscape. We acknowledge the hypothesis statement could be better, and have endeavoured to improve it as follows:

H3: Aggregating data to coarser spatial resolutions results in parameter estimates that increasingly fail to capture functional variations between land cover types, but stratifying the landscape prior to aggregation will reduce this functional information loss.