Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2022-109-RC1, 2022

Comment on bg-2022-109

Anonymous Referee #1

Referee comment on "Assimilation of multiple datasets results in large differences in regional- to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model" by Cédric Bacour et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2022-109-RC1, 2022

Summary

In this paper the authors report on a series of data assimilation experiments in which three different data streams (tower fluxes, MODIS NDVI, atm CO2 concentrations) were assimilated into the ORCHIDEE model individually, in pairs, or all together to assess the impact of multiple constraints on model parameters and predictions. The clearest take-away from the exercise was the importance of initializing soil carbon (C) pools by site or region to account for the non-equilibrium nature of the contemporary carbon cycle.

Major comments

In the Intro (L103-106) the authors raise the point that joint assimilation is "more optimal" than sequential assimilation but don't really explain the point, despite this being absolutely central to what is novel in this paper relative to previous ORCHIDEE DA papers. This really needs to occur, as from a theoretical perspective there should be NO difference between these two approaches – it's very easy to mathematically prove that under Bayes theorem, joint and sequential assimilations give the same answer. The authors acknowledge this somewhat in the Discussion, but even that doesn't really explain what's going on in their system. Given the numerical approaches used (numerical optimization instead of MCMC, SMC, or EnKF; covariances dropped or coarsely analytical approximated) it's unclear which results represent ORCHIDEE-specific problems and what

are general issues the rest of the community needs to worry about (and I'm somewhat inclined to think much of this is an artifact of how the ORCHIDEE sequential DA was implemented). To constructively move forward, please spend more time in the intro acknowledging and explaining these differences and more time in the Discussion reflecting on what the general take-home messages are and a bit of space on other Bayesian numerical methods and speculating about whether they are going to have the same issues.

Intro fails to include a clear statement of objectives, research questions, and hypotheses. The whole time I'm reading the rest of the paper I found myself asking "But why? What's the question you're trying to answer with this analysis?"

In terms of many of the high-level conclusions of the paper, it's not clear what here wasn't predetermined by structural choices of the Model Set-up (e.g. differences in which parameters were constrained depending on which data were assimilated). In particular the choice to spin-up carbon pools to equilibrium would automatically make it impossible for the model to match C observations unless C pool initial conditions were included in the calibration (the paper focuses on soil C, but undoubtedly this parameter is also indirectly absorbing some of the effects of the non-equilibrium vegetation C pools). This structural choice, more than anything about the DA itself or the choice of data streams or their information contribution, predetermines the main conclusion of the paper of the importance of assimilating soil C. Similarly, if you only let NDVI constrain phenology (not LAI, photosynthesis, etc) then it's obvious that NDVI isn't going to improve predictions of C fluxes or concentrations, and it's not surprising that it can actually make those things worse (since the default parameters will have compensating errors baked into them to get the right NEE with the wrong phenology, such that "fixing" the phenology will break NEE).

The set of assimilated data is notably out-of-date. The La Thuile flux data set (2006) was replaced by the FLUXNET2015 data set 7 years ago. The MODIS 5 data set (2005) is 3 versions behind the current collection (6.1, out since 2017), and the specific product used is considerably coarser than the actual data (0.72 degrees vs 250m) and then subsampled to just 15 pixels per PFT. In both cases the newer versions of NEE and NDVI are considerably larger in size that what was used and fix known errors in the previous versions. Similarly, the atmospheric CO2 constraint ends in 2009 and there is no mention anywhere in the entire paper of the satellite CO2 data sets that have absolutely revolutionized atmospheric inversions. There's a passing mention at the end of the paper (Line 867) that no data after 2010 is used in the analysis, but the authors don't say anything anywhere about WHY this is so and what impact this has on the inferences being

made. Why is so much data being left on the table? And if you're only going to use a subset of NDVI, why not include 250m data from the FLUXNET sites so that you gain the benefits of co-located constraints?

In the two step optimization, F+VI+CO2-2steps, it sounds like the CO2 data is being used twice, in both the first and second steps (i.e. double dipping). This is not OK and artificially inflates both overall sample sizes and the influence of the CO2 data. On a more minor point, how are you doing this without incurring all the disadvantages of sequential assimilation that are the point of the paper?

The approach used to specify the prior parameter uncertainties seems backwards and clearly represents double dipping (using the data being assimilated to specify the priors). Priors are supposed to represent the information you have about parameters BEFORE the data being assimilated is seen. I think if you want to use model outputs to put constraints on priors (e.g. the emergent constraints stuff Mat Williams has done with CARDAMOM) those constraints need to come from different data.

Results and Discussion section is WAY TOO LONG AND REPETITIVE. Section 4, which is just more Discussion, should be merged into the Discussion.

Specific points:

L34: "also given the technical challenges" feels tacked onto the end of the sentence. Not really explained and doesn't really provide any information. Either explain or drop

L97: This statement should acknowledge (cite and discuss) Trevor Keenan's "Rate my data" work, which took a really deep dive into the value of multiple constraints (albeit at a single site)

L107: on the issue of Likelihood weights, you should take a look at Oberpriller's paper http://doi.org/10.1111/ele.13728

L113: First, you raise the issue of systematic errors, but then the approach you use only accounts for random I.I.D Gaussian errors, so I'm not sure why you're bringing this up. Second, see papers by Istem Fer and Marcel Van Oijan (separately, not coathors) for examples of the formal accounting of systematic error during calibration

L165: Why was a linear relationship between NDVI and FAPAR assumed (and what was this calibrated against?). A quick google search turns up lots of papers that suggest a nonlinear, convex relationship between these two variables.

L198-207: how did you account for the uncertainties introduced via all these other data products?

L213: "locations are"

L253: misfit: this is more commonly called a cost function

L260-61: semantics note: observation error doesn't include model structural error, those are different concepts

L263: notational consistency: why would you abbreviate the prior as chi instead of writing this term out in quadratic form like all the other terms?

L272: missing a noun between "parameters" and "is calculated"

L273: Acronym TAF undefined

L344: The method described for estimating R sounds ad hoc and seems to constitute double dipping (using data that is later assimilated to set priors). More to the point, why not just estimate the observation error parameters are part of the calibration like any normal statistical model?

L349: I can't figure out what was actually done here based on this description

L395: (1) “large value of expresses a” seems to be missing a word or two. (2) a strong underestimation of observation error seems to be a problem that needs to be addressed more.

L417: In optimization, don’t we expect the norm of the gradient to be 0?

L422: Small notational preference: can you just call this RMSE like everyone else?

L435: This statement might benefit from adding qualifiers about this being difficult to estimate under the linear tangent optimization approach you are using. Other approaches to Bayesian computation don’t necessarily have this restriction.

L483: (1) undefined acronym. (2) This feels like the addition of new Methods in the Results – please go back and mention this in the Methods. (3) I think the paper would benefit from a figure that actually shows these seasonal cycles, and the model’s errors in replicating various parts of these cycles and trends, since they’re mentioned so often. I was hoping that was what Fig 2c was going to be, but not so much.

L529: Is this higher improvement due to the double dipping?

L550: "ORCHIDEE led with LMDz to overestimates…" I had to read this multiple times to figure out what you were trying to say. Even now, it only makes sense to me if I mentally cut out "with LMDz"

L556: undefined acronym

Pg 20: seems a bit bold to dive into the tropics vs mid-latitude debate given all the uncertainties in this analysis. But if you do, I'd recommend including Schimel et al 2015 https://doi.org/10.1073/pnas.14073021

L638: what do you mean that you include all the experiments in the GPP posterior? Each experiment should have it's own posterior and you haven't discussed any sort of Bayesian Model Averaging to be able to combine them

L799-802: Why is this not feasible? How would you make it more feasible?

L867: yes, but WHY?

L869: missing a word in “fluxes stocks”

Interesting that the Discussion doesn’t really mention how other teams are addressing some of the same problems you are struggling with (CARDAMOM, DART, PECAN, etc)

Figure 6 is completely illegible. This font size is WAY too small.