Reviewer #2

Overall response to both reviewers

We greatly thank both reviewers for the thorough and very helpful reviews. Synthesizing the two reviews indicated that manuscript had a lot of interesting information but was too dense to effectively communicate the key ideas. In response, we have simplified the analysis so that it has fewer moving parts. Our reanalysis also represents improvements to the data assimilation approach that have occurred since the manuscript was first submitted.

We simplified and modified the analysis as follows:

1) We removed the need for the 2-stage data assimilation. Now there are two chains that assimilate all sites simultaneously: one that includes site-specific parameters for only the Duke site and one that not include the site-specific parameters. This modification allows the analysis focus on why the parameters are different rather than focusing on the need to weight the Duke site differently. Since we did not actually weigh the Duke site differently in the original analysis nor include a synthetic experiment that explores the influence of site weighing on parameter inference, we feel that the simplified, updated approach is more sound and easier to understand.
2) We replaced the assimilations that separately removed the water and nutrient experiments with a single assimilation that removes all experiments (water, nutrient, and CO_2). We feel this is a better approach because the analysis included multi-factor experiments. For example, in the previous analysis, the removal of nutrient experiments also removed CO_2 and drought treatments. Now we present two sets of optimized parameters: with and without experiments. This allows us to more clearly address the question "how do the parameter distributions depend on the inclusion of ecosystem experiments in the data assimilation". Some of the figures were simplified in the process of this revision.
3) In response to Reviewer #2, we evaluated how well the model predicts the different experimental types. We now have a figure showing the observed and modeled experimental treatment responses for the data assimilation approaches. In the case of the data assimilation approach that did not include the experimental treatments, the comparison to the observed treatment responses are an independent validation of the model. We found that the data assimilation approach without the experiments predicts the experimental responses reasonably well, expect for the CO_2 experiment.
4) In response to Anthony Walker's helpful suggestion, we added an additional focus on regional predictions by simulating the regional response to nutrient addition, elevated CO_2, and drought. Our new analysis goes beyond the previous analysis by propagating the parameter uncertainty for all HUC12 units in the Southeastern U.S.
5) Our discussion section is re-worked to reflect the simplified analysis described above.
6) In response to comments by both reviewers to justify the set of parameters that were fit, we added six more parameters to the assimilation. We also removed the confusing reference to a sensitivity study of model parameters (the methods describing it were buried in the footnote of a table)
7) Sub-sections were added throughout to improve clarity.
8) We fixed some minor issues with the model structure as follows

a. The density independent mortality now removes entire individuals rather than the smallest individuals. This was accomplished by not using the parameter mS (the proportion of an average individual that is lost through turnover) in the density independent mortality calculation. Since density independent mortality represents random mortality it is more reasonable to not use mS in the calculation.
b. The model now simulates throughfall experiments directly rather than just reducing rain. Now rain is intercepted by the canopy in the full amount but the rain that enters the soil is reduced when simulating the throughfall experiment. This is a small change that makes the comparison cleaner.
c. FR is set to 1 in the fertilization studies that added nutrients at regular intervals. Many of these experiments were designed to fertilize to optimal nutrition so the assumption is well grounded and helps reduce the number of site level FR parameters that need to be optimized.
d. The process error terms are allowed to be a linear function of the prediction. This allows for the uncertainty to increase with the magnitude of the prediction. This linear function is applied to stem biomass, GEP, and ET. It allows for more confidence in predictions of lower values (like winter GEP and ET).

9) There were improvements to the data assimilation algorithm under the hood that allowed for faster run times and convergence. The cost function did not change (though we have described the cost function more clearly in the text).

Overall, the updated manuscript is more streamlined (though with more explanation in the methods section) and represents the state-of-the-art for the DAPPER algorithm.

Specific responses Reviewer #2
Our responses are in italics

Quinn Thomas et al. present a model-data fusion, or data assimilation, study that gathers 35 years of carbon cycle-related observations and manipulation experiments taken in Loblolly Pine ecosystems in the Southeastern US to optimize parameters of the 3-PG model within their new framework DAPPER. The authors examine the ability of the observations to constrain model parameters using a number of approaches for assimilating the different types of data, and they further examine the differences in model behavior/sensitivity and change in biomass stocks across the southeastern US as a result of the different experiments.

The authors have carried out an impressive and exhaustive collection of data for con- straining the 3-PG model in this study. This, and their investigation into different approaches for assimilating different types of data, in particular manipulation experimental data, make this study a noteworthy contribution to model–data assimilation literature in forested ecosystems, and therefore I would recommend publication in Biogeosciences. However, as it stands the manuscript is quite long and dense, which is understandable given the amount of detail that is required to present such a wide array of data and experiments. This being said, I recommend that the authors try to edit the article following some of the suggestions below (and their own views) to improve the clarity and readability of the text before this article is published.

- Overall, the objectives and key points of this study can get lost in the text. I think a few more sub-sections in the main text and supplementary, references and links between sections would help the reader to better follow and absorb the necessary amount of detail presented in the manuscript. I would also find it useful if the authors posed a few key scientific questions to help them highlight the main messages of the study.

  We clarified the last paragraph of the introduction to directly state the three objectives of the study. We also added section to the Methods, Results, and Discussion that parallel the objectives

- Some sections in the methods could do with more explanation for why certain approaches were used (see comments below) or better links to the supplementary material, as I have just mentioned.

  See comments below for response

- The introduction and discussion are quite long and this can prevent some of the key points from being highlighted. I suggest the authors try to cut down the text where they see fit, including some sentences that essentially are repetitions of earlier statements.

  We cut the introduction and removed paragraphs

- The paragraphs in the results section could be separate sections with sub-headings in order to guide the reader, while at the same time the results could benefit from stronger links between each section, especially before line 522, in particular comparing the between the 1st and 2nd stages, or the different 2-stage approaches with the 1- stage approach. At the moment, the results section before line 522 is a bit fragmented, making it harder to weave together a coherent story that brings out the key points.

  We added sub-sections to the results section

- Reading this manuscript I found myself asking: What do you expect from each experiment/approach? What will you gain/lose? Which approach is the right approach, going forward? These questions were largely answered in the discussion, and therefore I have made a suggestion below that perhaps some of the results and discussion could be merged within the sub-sections suggested above. This is a personal style issue however.

  We hope that updated analysis and discussion section helps answer these questions more clearly. There are now sub-sections in the results and discussion that help provide continuity between the sections

- Finally, the authors may consider cutting other sections of the discussion that are not fully pertinent to the results as the paper is already quite full of detail. I would like to stress that despite this suggestion I did find the discussion to be interesting and comprehensive, but I would like to see the key messages highlighted more and am concerned the length of the paper may overwhelm the reader.

We have cut out the paragraphs that aren't directly related to the results

Introduction

- Line 97: "relative contribution of each environmental control should be separated in order to correctly parameterize the sensitivity to changes in the environment". I agree to some extent but this is very hard to do and should we be separating each environmen- tal control, as the interaction between different environmental changes may produce different outcomes than if each were treated separately? I would be interested to hear the authors thoughts on this and what they think the impact of assimilating manipulation experiments data separately has on their results.

  Per reviewer #1 comment to shorten this paragraph, this sentence is now removed from the manuscript.

- Line 124-128: See previous studies Wutzler and Carvalhais (2014) and Section 2 of MacBean et al. (2016) for further discussion on debate of how to deal with the issue of weighting to account for the number of observations and/or using a multi-stage assimilation approach to address challenges of assimilating a diverse set of observations. Both issues are the subject of debate in the literature. On the issue of weighting by the number of observations, from a mathematical standpoint there would be no need if the error covariance matrix is properly characterized; however, this is difficult to achieve in practice. Similarly, a joint or simultaneous assimilation, in which all observations are assimilated together, is mathematically more rigorous as the error covariance between the observations can be properly taken into account. I appreciate that you have dis- cussed the benefit of weighting by the type of data in the discussion, but this debate in the literature (for and against weighting, due to the abovementioned reasons) should perhaps be referred to more clearly in this study.

  Per reviewer #1 comment to shorten this paragraph and review #2 comment that the discussion lacks of the data weighting lacks precision, we cut this discussion.

- Line 129: It is true of course that to constrain changes in biomass monthly time-scale models are sufficient, but note that monthly time-scale models are not the only way to overcome computational challenges associated with inverting a complex ecosystem model. There are sophisticated yet simple algorithms that dramatically improve the sampling of parameter space in a limited number of iterations. See the work of Jasper Vrugt: https://scholar.google.co.uk/citations?user=zkNXecUAAAAJ&hl=en&oi=ao

  We cut the discussion about monthly time-step models while shortening the paragraph but will definitely look more closely into the work by Vrught. Thanks for highlighting!

Methods

- Section 2.1 It would be good if you could refer to references and/or relevant sections in the Supplement in Section2.1 to depict between standard characteristics of the 3PG

model specific additions or alternative choices you made and (and to explain why you made those choices). For example:

Added subsections to the Supplemental Material and added the references to Supplemental Material to the main text

- Line 201-202: Was this additional function based on a published study?

  The function was developed as part of this study

- Line 209: Is the site-index a new addition to the model that you developed? If so, from where?

  The text now reads:
  For unfertilized plots, we used site index (SI), a measure of the height of a stand at a specified age (25 years), to estimate FR. This approach is in keeping with previous efforts (Gonzalez-Benecke et al., 2016; Subedi et al., 2015)

- Lines 218-220: Why did you remove the dependence of total root allocation on FR for the DA study?

  We removed the dependence of total root allocation on FR because we separated root allocation into the coarse and fine roots. Therefore, the previous function was not applicable. Future studies should investigate how best to build this function back in and ask whether we currently have the observational constraints to parameterize it.

- Line 229-231: A reference for or further explanation of this modification would be good here. –

  Added text

- Line 245: "implicit irrigation in very dry conditions." Is this a realistic feature of these sites? How does this affect the results? Especially for the water availability manipulation experiments.

  We added text explain how this assumption could influence the results. "This assumption may cause the model to be less sensitive to low soil availability but the optimized parameterization may compensate. "

- Line 250: do you mean to say "mean monthly GPP"?

  GPP was a sum for each month so 'monthly GPP' is correct. Mean monthly GPP might imply that multiple months are averaged.

- Line 251-252: How did you select the 31 parameters to be optimized?

In the revised manuscript, we included more parameters that were optimized (six more). The eight parameters that were not optimized did not have specific data to use as a constraint (leaf boundary layer, conductance, canopy light extinction coefficient, etc).

- Table 1: Please can you give the equation for how the sensitivity is calculated? Also, please could you explain why there is both a number and "vague" given for the uncer tainty of some parameters? If "vague", please can you detail how you defined the prior uncertainty/ranges in the text?

  We cut out the sensitivity analysis and added more parameters to the optimization.

- Finally, I appreciate you have a lot of information to con- vey and the tables are large, but it might be good to have all optimized parameters here and just indicated which ones are referred to in the discussion.

  We expanded the table to include all optimized parameters

- As a general comment, it is hard to find some of the information you refer to in the Supplement (e.g. the other optimized parameters you refer to in the caption of Table 1). Please could you split the Supplement into numbered/indexed sections and then refer specifically to the relevant section to help the reader?

  We added section divisions to the supplemental material

- Line 255-265: How did you initiate the biomass pools? Based on site-level data for the start of the simulation period? Please detail with references. If no site data were available, how sensitive were your DA experiments on the method used to initiate the biomass pools? Later note: I see you have addressed this in Section 2.4. It might be useful to refer to that section here so the reader is not questioning this in this section.

  We moved the text on the initialization described to the section on the model description

- Section 2.2 Table 2: Last column – Table 3 instead of Table 4. Also, please could uou explain, or give references, for why the SD for observations sometimes varied between 10% and 2.5% of the observation.

  To reduce confusion, we used 10% for LAI observations. Future applications of the method can focus more on the influence of data uncertainty on parameter estimates.

- Section 2.3 Equation 4: Please explain why you picked a uniform distribution between 0.001 and 100?

We added text to state that the bounds of 0.001 to 100 were designed allow the priors to be vague. The bounds include reasonable ranges of standard deviation parameters.

- Lines 348-349: Please explain why (only) 3 MCMC chains were run? Was a convergence metric such as R-hat used?

  We re-ran our optimization with the updates described at the top of the response. We ran 4 chains and used the Gelman R criteria to test for convergence.

- Section 2.4 Lines 398-399: Although I understand the reasoning that these sites are close together and the most data rich, I don't understand why you lump the Duke CO2 enrichment site with DK3 and NC2 in the 1st stage when you stated that you wanted to test the influence of the CO2 fertilization – why not just test the Duke CO2 enrichment site by itself in the 1st stage and the remaining sites/plots in the 2nd stage to answer this question?

  Our updated analysis removed the need for a 2-step analysis. (see beginning of this response for more info)

- Further to the above point, I appreciate the extra experiments to understand the influence of the CO2 fertilization on the posterior parameters, and the further experiments to determine the influence of the water treatments and nutrient addition. But how dependent are your results on which type of observation and/or treatment is assimilated in the 1st stage vs 2nd stage? Would the results different if you reverse the stages you have in your current set-up? Again, see Wutzler and Carvalhais (2014) and/or MacBean et al. (2016) who discuss these issues (as well as the issue of the weight of different types of data, as you discuss below. A pseudo-test with synthetic observations would have been useful prior to assimilating real data to determine whether the exact set-up of a 2-stage assimilation is sensitive to the order of observation assimilation as well as to confirm if the assimilation system is able to constrain the parameters to their correct values.

  Our updated analysis removed the need for a 2-step analysis. (see beginning of this response for more info)

- Lines 430-465: While the tests and approaches put forward here are interesting, the text is dense. Any efforts the authors could make to simplify the description of the experiments and simulations performed (perhaps with the use of a table and simula- tion/experiment code names?) would likely help the reader.

  We reorganized and clarified this text in response to this comment and comments from Reviewer 1

- Lines 467-475: The cross-validation exercise presented here is a useful one. Was a similar test used to assess the validity of the posterior distributions of the manipulation experiments, even though there are fewer sites?

We added a cross validation of the experiments treatments. We now include optimized parameter set that did not include the experimentally treated plots in the assimilation. This parameter set is now used to predict the experimental treatments.

Results

- Line 480-484: Description of the sensitivity analysis and choice of parameters should be in the methods. Was this a one-at-a-time sensitivity analysis or a full global method? What is the justification for using this approach versus an existing global sensitivity analysis that accounts for correlations between parameters and explores the whole parameter space (unless I have misunderstood what was done)?

  We cut the reference to the sensitivity analysis

- Why did you fix the light extinction coefficient as opposed to the quantum yield parameter?

  We fixed the light extinction coefficient because it was more known than the canopy quantum yield.

- Supplemental Table 3 and Table 5: As mentioned above I would suggest having all the optimized parameters in one table. I would also suggest putting the prior min/max in Table 5 even though it might mean having an extra line/column per parameter and taking this information out of table 1 so it is easier to see how well the optimization has constrained the parameters.

  We moved all parameters to the table in the main text and added the range uncertainty in the priors to the same table

- Finally, I would suggest splitting up the parameter tables into the sections you refer to in the text, e.g. "temperature sensitivity of quantum yield" or "physiological parameters" etc. This will make it easier for the reader to refer to the tables when reading the text.

  Done

- Which experiment do the supplemental figures correspond to? The "ALL" experiment? This should be detailed.

  The assimilation approaches have been renamed and clarified in the supplemental figures.

- Are you talking about the 1st stage experiment in the first paragraph of the results? If so, it would be good to specify this, and I would further suggest splitting the results into sections to more easily guide the reader.

We clarified by using the names of the data assimilation approaches. Our results section is better organized in response to review 2.

- Do you discuss DK+NC2-fert in the results, or have I missed it? Perhaps more needed on the 1-stage versus 1st and 2nd stages before you discuss the experiments with and without nutrient and water addition (i.e. before line 522)?

  Our updated analysis did not require the 2-stage approach so we no longer need to report the DK+NC2-fert results

- Figure 5 comes before Figure 4 in the text – switch around?

  Fixed in text

- Lines 507-515: I am a bit confused by the sentence "The two-stage assimilation was critical for constraining the CO2 quantum yield enhancement parameter (Calpha700)" as you then go on to say (and show, in Figure 5) that the 1 stage resulted in a narrower uncertainty interval? I guess you mean that despite the higher 95% confidence interval, the 2-stage approach results in a more realistic parameter value but I am not at all sure on that? Please could you clarify this in the text?

  Paragraph was modified in the revisions

- Line 517: I would suggest putting the names of the soil fertility parameters in brackets to aid the reader, or again put sub-headings in the parameter tables.

  Paragraph was removed during revisions

- As you did not have a strong difference in predictive capability between experiments with and without nutrient or water addition, even though you had different parameters, that presumably means you have a certain amount of model equifinality? You discuss and show the difference in model behavior as a result of the different approaches in Figures 5 – 7, but you do not discuss which one you think leads to the right behav- ior? Do you have an idea? Perhaps a synthetic experiment with pseudo-observations taken from the model simulations might help with this (a so-called "observing system simulation experiment", or OSSE)?

  This was a very insightful comment. Our response reflects the updated analysis described above that has two assimilation approaches: with and without ecosystem experiments. Our new Figure 5 (the bar graph with the experimental responses from the observations and model predictions) helps support the following:
  - *Including experiment substantially increases the predictive capacity of the model in the CO_2 experiments.*
  - *The predictive capacity of drought, irrigation, and nutrient fertilization experiments did not substantially change whether experiments where included or not.*

We think that an OSSE would be a great follow on study that more specifically explores of the issues that are brought up in this analysis. An OSSE could explore how locations of plots within a region and the different types of individual experiments influence the ability to retrieve known parameters. Such a study would build on the description of the cost function and general approach presented in this manuscript. Since we do not include an OSSE, we now try to avoid making general statements in the discussion that would require an OSSE to quantitatively support.

- Lines 522 onwards show very interesting results. However, I would suggest that the patterns detailed in last two paragraphs (Lines 553-572) would benefit from explanations linking back a bit more (not just referring to figures) to the different model behavior/mechanisms identified and discussed in the RW-fert and RW-water sections just above.

  In response to Reviewer 1, we cut the results of the regional simulations from the RW-fert and RW-water simulations

Discussion

- First paragraph is more of a summary than a discussion and could be cut or added to conclusions.

  We prefer provide a summary at the beginning of discussions to remind the reviewer of key points.

- Although perhaps a little too long, this is a useful discussion that ties the results to- gether and answers some of the questions I raised in my comments on the results. Perhaps it would be useful to combine some of the summary points raised in the dis- cussion with relevant sections in the results with separate sub-headings as I mentioned above.

  We added subheadings to the discussion

- Lines 650-652: Interesting point and in addition, as I have mentioned above, I think a synthetic experiment would also be very helpful in this regard.

  We agree that a synthetic experiment would be an excellent next study. The synthetic experiment could create 'fake' region with different environmental gradients and explore the types of gradients that allow for the retrieval of parameters from the OSSE study.

Minor comments

- Line 87: Do you mean the "assimilation of manipulation experimental data", rather than the "assimilation of experiments"?

  Yes. Fixed

- Line 88: two or more

  Fixed

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., and Schürmann, G.: Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, Geosci. Model Dev., 9, 3569-3588, doi:10.5194/gmd-9-3569-2016, 2016