Reviewer #1 (Walker)

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 #1 (Walker) below
Our responses are in italics

Thomas et al present a data-assimilation (DA) study using constraints from multiple data streams from multiple sites and experiments to optimise parameters in the monthly timestep PG-3 model of loblolly pine production. The study has three specific objec- tives. Stated on lns 170-171, 1) a new regional and hierarchical data assimilation sys- tem with the capacity to assimilate multiple data streams from multiple experiments; stated on ln 179-180, 2) the consequences for parameter estimation and prediction of including or not including ecosystem manipulation experiments (this could be more broadly stated as evaluation of the DA); and stated on ln 181 3) model predictions with the optimised parameter set of forest biomass changes in response to changes in nutrient addition of precipitation. This study is well thought out and implemented, presents a useful advance to the use of DA in ecosystem modelling and forecasting, and will likely be of interest to many readers of Biogeosciences.

My main criticism is that the distinction between the three areas of this study is often not made explicitly throughout the manuscript and consequently the manuscript is not as readable or as clear as it could be.

The majority of my comments are an attempt to help improve the organisation and presentation of the manuscript with the goal that this study will be as widely read and cited as possible.

- With that in mind, I suggest organising the manuscript as much as possible by the three stated objectives. I suggest combining the sentence on lns 179-180 with the sentence on lns 170-171 and explicitly listing the three objectives together. The results and discussion section would benefit from organisation along the lines of the three stated objectives. I suggest breaking each into three subsections, each dealing with one of the objectives. Again the conclusions section should specifically address each objective.

  We have modified the structure of the manuscript so that the methods, results, and conclusion now have sections that address each of the three objectives.

Abstract

- It would be good to be specific about who the target audience is for this research. The research straddles a technical field that develops DA but the technique produces a tool at a level of maturity that could be used by foresters. These ultimate end users could be more explicitly targeted.

  Text added

Introduction

- Is a bit long and could a page or so could be cut without loss of content. Paragraphs on lns 82-105 could be combined and reduced in length. The main point is that ecosystem experiments can help to reduce the problem of equifinality in DA.

  The paragraphs between 82 and 105 were shortened

- The paragraph on lns 108-141 makes some nice points but could be substantially shortened without loss of content. Much of the paragraph is methods like.

  We kept the first sentences of this paragraph that highlight the value of regional data-assimilation but, per the recommendation, we cut the remaining paragraph

- Weight to rare experiments (mentioned on ln 125) could also apply to rare data types. Later in the paragraph (ln 135-136) the authors state that data of different frequency is a problem in biasing the cost function toward high frequency data, but offer no solution other than a monthly timestep model. Rare data, or low frequency data, could also be given higher weights. Also high frequency data could be summarised at lower frequency.

  The discussion of the data weighting was removed to shorten and tighten the paragraph.

Methods

Again long and could probably be made more concise. Also the organisation is tough to follow.

- I suggest leading with the observations, the various sites, and measurement campaigns/projects.Many of these are not properly introduced. This will provide a comprehensive introduction to the system and what measurements actually go into this DA system. Observation sites and projects are mentioned on ln 409-410, but these are not introduced and need to be described in the observations section of the methods.

  We moved the section on the observations to the beginning of the methods section. We structured the paragraph so that it gives an overview of all the measurement campaigns.

- I found section 2.3 very difficult to follow. I'm not expert on DA mathematical meth- ods but I have a reasonable conceptual handle on DA, and yet I was lost in the first paragraph. I also ran this section by a colleague who is expert in the mathematics underpinning DA and they agreed that this sections needs to be clearer. Their key criticism was that they could not see the derivation of Eq 7, perhaps the authors could add the derivation to an appendix. And that it is not clear how the MCMC was used to sample Eq 7. A clear description of the details of the MCMC procedure is necessary, along with the presentation of the cost function. Also the first term on the righthand side of Eq 7 is not the same as the righthand side of Eq 1, is this deliberate? And E is never defined.

  I strongly suggest reworking section 2.3 of the methods to be extremely clear about the DA process and how it was implemented. Start with a clear description of the goals of the DA – state estimation and estimation of parameter distributions. Then describe all the various sources of uncertainty and how the method accounts for them. Then take the reader step by step through the method. Perhaps a diagram would be useful. The following comments are an attempt to provide examples of where confusion arises but they are in no way comprehensive. The sentence on lns 281-283 is more or less stating the the same thing as the sentence on lns 284-285. I suggest fusing these together. Is the reference to a "latent model" really necessary, it is confusing with the mathematical model. Would "true" system states and fluxes convey the same meaning? Do not try to justify the method in comparison with previous methods (e.g. lns 286-291), in the methods this just confuses the description and this can be argued in the discussion. On lns 291-293, this is state estimation right? That's fine but is it really the focus of your method? None of the three stated objectives are for state estimation. How exactly was estimation of the latent state or flux the first step in the process when it includes the optimised parameters etc as described on lns 296-298? Seems like the statement on ln 306-308 should come before the previous paragraph.

  We cleaned up the description of the cost function per the reviewer recommendation. (see Supplement to the review)

We used a hierarchal Bayesian framework to estimate the posterior distributions of parameters, latent states of stocks and fluxes, and process uncertainty parameters. The latent states represented a value of the stock or flux before uncertainty was added through measurement. The approach was as follows.

Consider a stock or flux (m) for a single plot (p) at time t ($q_{p,m,t}$). $q_{p,m,t}$ is influenced by the processes represented in the 3-PG model and a normally distributed model process error term,

$$q_{p,m,t} \sim N(f(\boldsymbol{\theta}, FR_p), \sigma_m) \qquad \textit{Equation 1}$$

where $\boldsymbol{\theta}$ is a vector of parameters that are optimized, FR_p is the site fertility, and σ_m is the model process error. Not shown are the vector of parameters that were not optimized (Supplemental Material Table 1), the plot ASW, an array climate inputs, and the initial conditions because these are assumed known and not estimated in the hierarchical model. The process error assumed that the error linearly scales with the magnitude of the prediction:

$$\sigma_m^2 = \gamma_m + \rho_m f(\boldsymbol{\theta}, FR_p) \qquad \textit{Equation 2}$$

While the structure of the Bayesian model allowed for all data streams to have process uncertainty that scales with the prediction, in this application we only allowed stem biomass, GEP, and ET process uncertainty to scale because they had large variation across space (stem biomass) and through time (i.e., there should be lower process uncertainty in the winter when GEP is lower). For the other data streams, the linear scaling term was removed by fixing ρ_m at 0.

FR_p did not have an explicit probability distribution. Rather the probability density evaluated to 1 if the plot was not fertilized, thus causing FR_p to be estimated from SI and MAT (Supplemental Material Equation 15), or if it was a fertilized plot and has an FR_p equal or higher than that of its non-fertilized control plot. The probability density evaluated to 0 if the estimated FR_p in a fertilized plot was less than the FR_p in the control plot or FR_p awas not contained in the interval between 0 and 1.

$$FR_p \sim \begin{cases} 1 \text{ if non-fertilized, } FR_p \geq 0, \text{ and } FR_p \leq 1 \\ 1 \text{ if } FR_p = 1 \text{ and fertilization levels are assumed to remove nutrient deficiencies} \\ 0 \text{ if } FR_p < 1 \text{ and fertilization levels are assumed to remove nutrient deficiencies} \\ 1 \text{ if fertilized but levels are not assumed to remove deficiencies and } FR_p \geq FR \text{ of control plot} \\ 0 \text{ if fertilized but levels are not assumed to remove deficiencies and } FR_p < FR \text{ of control plot} \\ 0 \text{ if } FR_p < 0 \text{ or } FR_p > 1 \end{cases}$$

Equation 3

Our model included the effect of observational errors for measurements of stocks and fluxes. For a single stocks or flux for a plot at time t there is an observation ($y_{p,m,t}$). The normally distributed observation error model was:

$y_{p,m,t} \sim N(q_{p,m,t}, \tau^2_{p,m,t})$ *Equation 4*

where $\tau^2_{p,m,t}$ represented the measurement error of the observed state or flux. By including the observational error model, $q_{p,m,t}$ represented the latent, or unobserved, stock or flux. The variance was unique to each observation because it was represented as a proportion of the observed value. The $\tau^2_{p,m,t}$ was assumed known (see Table 2) and not estimated in the hierarchical model (Table 2).

The hierarchical model required prior distributions for all optimized parameters, including the parameters for the 3-PG model (θ), FR_p, and the process error parameters. The prior distributions for θ are specified in Table 3. Some parameters were informed by previous research in loblolly pine ecosystems while other parameters were 'non-informative' with flat distributions (termed 'vague' in Table 3). The prior distributions for the process error parameters were non-informative and had a uniform distribution with upper and lower bounds that spanned the range of reasonable error terms.

$\gamma_m \sim U(0.001, 100)$ *Equation 5*

$\rho_m \sim U(0, 10)$ *Equation 6*

By combining the data, process, and prior models, our joint posterior that includes all thirteen data streams, plots, months with observations, and fitted parameters was

$$p(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{q}|\boldsymbol{y}, \boldsymbol{\tau}, priors) \propto$$

$$\prod_{p=1}^{P}\prod_{m=1}^{M}\prod_{t=1}^{T} N(q_{p,m,t}|f(\boldsymbol{\theta}, FR_p), \gamma_m + \rho_m f(\boldsymbol{\theta}, FR_p))$$

$$\prod_{p=1}^{P}\prod_{m=1}^{M}\prod_{t=1}^{T} N(y_{p,m,t}|q_{p,m,t}, \tau^2_{p,m,t})$$

$$\prod_{p=1}^{P} p(FR_p) \prod_{f=1}^{F} p(\theta_f) \prod_{m=1}^{M} p(\gamma_m) \prod_{m=1}^{M} p(\rho_m)$$

Equation 7

where bolded components represent vectors, P is the total number of plots, M is the total number of data streams, T is the total months with observations, and F is the total number of 3-PG parameters that are optimized.

We numerically estimated the joint posterior distribution using the Monte-Carlo Markov Chain – Metropolis Hasting (MCMC-MH) algorithm (Zobitz et al., 2011). This approach has been widely used to approximate parameter distributions in ecosystem DA research (Fox et al., 2009; Trudinger et al., 2007; Williams et al., 2005; Zobitz et al., 2011). Briefly, the algorithm proposes new values for the model parameters, uncertainty parameters, latent states, and FR. The proposed values were generated using a random draw from a normal distribution with a mean equal to the previously accepted value for that parameter and standard deviation equal to the parameter-specific jumping size. The ratio of proposed calculation of Equation 7 to the previously accepted calculation of Equation 7 was used to determine if the proposed parameters are accepted. If the ratio was greater than or equal to 1 the proposed values were always accepted. If the ratio was less than 1, a random number between 0 and 1 was drawn and the proposed values are accepted if the ratio was greater than the random number. This allowed less probable parameter sets to be accepted, thus sampling the posterior distribution. We adapted the size of the jump size for each parameter to ensure the acceptance rate of the parameter set was between 22% and 43% (Ziehn et al., 2012) by adjusting the jump size if the acceptance rate for a parameter is outside the 22 – 43% range. All MCMC-MH chains were run for 30 million iterations with the first 15 million iterations discarded as the burn-in. Four chains were run and tested for convergence using the Gelman–Rubin convergence criterion, where a value for the criterion less than 1.1 indicated an acceptable level of convergence. We sampled every 1000th parameter in the final 15 million iterations of the MCMC-MH chain and used this thinned chain in the analysis described below. The 3-PG model and MCMC-MH algorithm were

programed in FORTRAN 90 and used OpenMP to parallelize the simulation of each plot within an iteration of the MCMC-MH algorithm.

- Section 2.4 jumps around between objectives. Some text would fit better in section 2.3, for example lns 408-428. Text on lns 454-461 would be better organised if it were to follow the text on 430-444, then the regional simulations can be presented afterwards.

  We reorganized as suggested by the reviewer

- I suggest defining sections 2.3, 2.4, and an additional 2.5 to be organised by the three stated objectives.

  We reorganized as suggested by the reviewer

- Also, while commonly used by the modelling community, I do not agree that you can run "experiments" with models. Models make predictions from a specific set of mathematical hypotheses and defined scenarios. An experiment is designed to test predictions and discriminate among hypotheses.

  We removed the 'experiments' language

Results

- Why were only 31 parameters optimised, can you describe why this set were chosen from the total 46?

  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).

- Technically the parameters are not "sensitive" (ln 480), it is the model output that is sensitive to the parameter. "Influential" would be a better adjective to describe the parameters.

  To simplify the analysis and reduce the density of the manuscript we removed the sensitivity study and the reference to it in the text.

- Lns 486 & 488 variability is described as being reduced but no data are provided. Can you quantify these statements. There are many statements like this throughout the results and they ought to be quantified (e.g. lns 502, 508). Also on 508, is mean correct, isn't this the median of the parameter distribution?

We added a column to the table that is the ratio of the size of the posterior 99% credible interval to the size of the prior 99% confidence interval. This ratio illustrates how the uncertainty is reduced by the data assimilation.

- Some kind of visual representation of the data in table 5 would be useful.

  Supplemental Material Figure 1 shows the PDF of the prior and posterior

- Ln 492 what do you mean strong priors? Well defined from measurements and litera- ture with low variance? Could you quantify this?

  We removed this language from the manuscript to reduce confusion

- Lns 494 the process uncertainty parameters are mentioned here and in the methods, but results are barely presented (only in the supplement) and are not discussed, or not that I noticed. This is a very interesting concept and I would like to see these data pre- sented a little more and at least a little discussed. What kind of impact does including these parameters have on the optimised parameter distributions? I understand you are already presenting a lot, but this is fairly novel as far as I'm aware and is of interest.

  We added a small discussion of the process error parameters to the discussion section

- Figure 10 and 11 would be more in keeping with your stated goal of forecasting on lns 65-68 if you removed the b panels in both plots. If you think that the parameter estimates when including the data from the manipulations gives a better estimate of those parameters then the data in panels b are not particularly useful for forecasts. In my view, and as stated on line 67 & 68 "provide information on both the expected future state of the forest and the probability distribution of those future states", the final figures would be much stronger if the probability distribution of the future states shown on the a panels were represented on the b panels.

  We combined the Figure 10 and 11 into a single figure that has the median prediction on the left side and the uncertainty on the right side. This allows the figures to represent the forecasting capacities of the data assimilation approach.

- While it is interesting to show the consequences for prediction of inclusion of manipulations or not, and the opposite sign of the change in predictions when water and nutrient manipulations are included, you already show this in Figures 6 & 7. If you want to keep the b panels in 10 and 11 I suggest you add them as extra panels to figures 6 & 7, showing the absolute delta (or similar) from the simulations that include the manipulation delta. This will allow you to address the question: what are the consequences of not including data from manipulations? Without confounding the predictions from the most appropriate DA product for the scenarios tested. Also, the scale ought to be the same for the data presented in Figs 10 and 11.

  We cut panel b from these figures (see above)

- Was CO2 change included in the above projections of removal of nutrient limitation and precipitation reduction? Furthermore, it seems you have included data from water manipulation experiments, nutrient manipulation experiments, and CO2 manipulation experiments. But you have only made projections for nutrient and precipitation change. Why not CO2 change? CO2 projections would complete the study.

  We added a +200 ppm simulation to the set of regional predictions. The predicted regional changes are for +200 ppm, -30% precipitation, and removal of nutrient limitation. The uncertainty for each prediction is shown.

Additional points

- I think the title would benefit from the addition of "Loblolly Pine".

  Added to title

- Ln 50 Duke FACE experiment had 4 replicate plots, so where does the 5 come from on this line. An additional plot from the unreplicated prototype?

  We removed the language from the abstract and later in the text we clarified that the replicated prototype was used (per the data reported in McCarthy et al. 2010)

- Ln 48 – 50 the sentence on this line would help flow if it were before the preceding sentence.

  Revised

- Ln 65 I don't think I would classify the three areas mentioned in the previous sentence as tools. They are more than tools, they are also knowledge.

  Removed the word 'tools' so that the sentence references the previous sentence terminology ('sources of information')

- Ln 67 What do you mean by "based on" here. Can probably delete. Also while I think your methods could be used for "forecasting" you don't really use the method in that sense.

  Removed the clause that contained 'based on'

- Ln 73 insert "can" in between "that generate"

  Fixed in text

- Ln 85 86 "carbon allocation and turnover" This is worded a little awkwardly

  Removed awkward language from text

- Ln 97-99 awkward way to start a paragraph.

  Paragraph was removed during the shortening of the introduction

- Ln 111 suggest replacing "important" with "useful" or something more descriptive

  Changed to 'useful'

- Ln 155-157 suggest replacing "nutrients" with "nutrient addition". Also suggest removing hyphens.

  Changed in text

- Ln 162-163 Awkward

  Removed 'available' to make less awkward

- Ln 171 Again I think you need to call out loblolly pine here

  Changed in text when revision the statement of objectives

- Ln 175 The authors chosen acronym, in my view, somewhat undersells what they are doing. The DA method is hierarchical and considers data from multiple sites and of multiple different types. The acronym gives not indication of this and suggests that the DA method is only suitable for Pine Plantations. Of course it is the authors' choice though.

  Thank you for the suggestion to broaden the acronym. We kept the same acronym but changed the words to "Data Assimilation to Predict Productivity for Ecosystems and Regions" to emphasize the multi-site aspect of the DA.

- Ln 307 insert "considered" between "was a"

  Sentence was modified during revisions

- Ln 446 replace "regional" with "region"

  Changed in text

- Ln 522-524 I'm not sure what you mean here, could you clarify?

  Sentence removed during the revisions

- Ln 528 delete "a"

  Done

- Ln 576 replace “detangling” with “disentangling”

  Done

- Ln 582 I think “synthesised” would be a better word to use than “organised”

  Done

- Ln 591-591 I take your point about equifinality but can you really say this if predictions were not improved in some way? Just a thought. Is there a way that you can be sure that the mechanisms were correctly distinguished?

  We removed this sentence during revisions

- Ln 633-634 Agreed, but did your method strictly weight the data? Wasn’t it more that the hierarchical method gave priority to the CO2 manipulation data?

  We removed this sentence during revisions

- Ln 646 replace “than” with “that”

  Done

- Ln 656 quantify this statement

  We removed this sentence during revisions

- Ln 662-663 this was news to me when I read this sentence. I think this would become clearer once the methods can be clarified as suggested above.

  We clarified in the method section. The method section more completely describes assumptions of the site index estimation.

- Ln 668 suggest changing “prior” to “previous”, just to maintain the meaning of prior in the Bayesian sense.

  Done

- Ln 673 you do not show any data on covariation of parameters.

  We removed this language

- Ln 676-680 I like this statement, makes a lot of sense. But is it most appropriate here? This point should be made clearly in the methods.

  Moved to methods

- Ln 685 suggest deleting “Multivariate Constructed Analogs (MACA)” it is not needed.

  Deleted

- Ln 692-697 This is a good point but I’m curious why the change in biomass in response to precipitation reduction was small given the large change in parameter values when water manipulations were included in the DA. Can you try to explain this based on the process hypotheses embedded in the model.

  We cut this sentence during revisions

- Ln 698 replace “reduced” with “reduction” Ln 707 insert “as a function of”

  Done

- Ln 719 insert space in “fromadditional”

  Done

- Ln 760 While I’m sure the methods and tools developed by this study could be used for ecological forecasting, strictly speaking this study is not ecological forecasting. The third objective, which concerns optimised model predictions, is a scenario analysis rather than a forecast.

  We removed the term ‘ecological forecast’ from the sentence

- Ln 769 no need to cite Medlyn et al 2015 here

  Removed citation