Biogeosciences Discuss., 11, C6518–C6526, 2014
www.biogeosciences-discuss.net/11/C6518/2014/

Biogeosciences Discussions
Open Access

Interactive comment on "Meta-analysis of high-latitude nitrogen-addition and warming studies imply ecological mechanisms overlooked by land models" *by* N. J. Bouskill et al.

N. J. Bouskill et al.

njbouskill@lbl.gov

Received and published: 7 November 2014

We thank the reviewer for their constructive comments on our manuscript, they have been very helpful for re-focusing the manuscript, and drawing out some of our ideas for how a data set like this can highlight appropriate and inappropriate benchmarks for testing the models.

Below are the different comments from the reviewer followed by our responses.

The main strength of the paper is that it presents both a new meta-analysis of high latitude warming/N additions experiments and a model-data comparison. It uses the results to demonstrate key patterns that fundamentally differ between the models and the field studies. The largest weaknesses of the paper are the long and challenging to follow discussion and the lack of key information for the simulation protocol that could influence the results.

The discussion and conclusion section read like the authors are laying every issue with the CLM on the table. The manuscript could greatly benefit from a better-organized discussion that clearly distinguishes the important points from the secondary points. Furthermore, section 4.3 seems to be about issues associated with the model-data comparison, but most of the paragraphs in the section don't address issues with benchmarking. For example, the paragraph on nitrogen fixation only addresses issues with nitrogen fixation not benchmarking. I recommend focusing this section.

We agree with the reviewer that the manuscript discussion was too long have revised it in the current draft of the manuscript, reducing the length by 2 pages and introducing more sub-headings to help with flow. Furthermore, we have restructured aspects in accordance with the reviewer's comments below.

Individual scientific/issues

Page 12377, Line 14: I would emphasize the role of nutrients in climate-decomposition increased growth rates feedbacks. As it reads, decomposition directly increases growth rather than increases in N mineralization increasing growth.

We have re-written this section for clarity.

Page 12377, line 26 – Page 12378, line 10: The argument for why a meta-analysis approach is different from the site-level comparison used in Thomas et al. 2013 is not clear. Both approaches use perturbation simulations in ESMs and extract gridcell level output that correspond to grid-cells with experiences. Both approaches use short-term perturbations. Both approaches focus on means across many sites. Is the unique contribution the use of meta-analysis statistics? It seems that the Thomas et al. study is broadly similar to this study and the key distinction being drawn here is between model-

BGD

11, C6518–C6526, 2014

data comparisons that use many sites and focus on broad patterns but don't try to simulate each site perfectly and model-data comparisons that focus on a few sites but focus on matching the conditions of the study perfectly and examine detailed dynamics. (e.g., the FACE comparison by Zaehle et al. 2014). Overall, a better argument for why the meta-analysis approach is unique and particularly useful is needed.

It was not our original intent to directly compare our approach with that laid out in Thomas et al., 2013. However, we realize that it is easy to interpret this paragraph in this way. We believe the approach described by Thomas et al., 2013 and the approach laid out in the current manuscript are largely complementary. The point we were trying to make is that such an approach doesn't capture the spatial heterogeneity of responses the way a data synthesis of studies spanning thousands of kilometers might. We have re-phrased this paragraph to avoid confusion.

Page 12379, Line 7: How was GPP estimated? Was GPP a modeled outcome from the partitioning of NEE into GEP and RE? If so, this should be stated.

This line actually refers to the measured GPP from the field experiments. However, the modeled GPP was simulated directly from CLM as net leaf photosynthesis using the Farquhar model (Farquhar et al. 1980) for C3 plants and the Collatz model (Collatz et al., 1992) for C4 plants.

Page 12379, Line: It might be useful to list the summaries statistics (range and mean) for the warming in the observations. It would help the reader understand why âĞă1C warming was targeted in the CLM simulations

This information is given in the results, however, we have added a sentence directing the reader to these data.

Page 12379, Line 20: The focus of the manuscript is on nitrogen-carbon interactions but studies with P and K were used. How many studies were multi-element additions? How would this influence the results?

BGD

11, C6518–C6526, 2014

In the present data set approximately 0 to 40% (depending on the response measured) of the studies used an NPK fertilizer instead of NH4NO3, and these studies were generally in Europe. Table S1 in the supplemental material breaks the responses down by N-species added and there is an affect on microbial biomass from using NPK compared with NH4NO3. However, other response ratios (e.g., below ground respiration) came from studies that only used NH4NO3 as the nitrogen source and in general we included only data using NH4NO3- fertilization, so we don't believe the use of NPK has influenced the overall conclusions to a great degree.

Page 12382, first paragraph in section 2.3: Model protocol description is severely lacking. For example, what resolution was the model run? What climate forcing was used? Was 1500 years suitable for the carbon stocks to come to equilibrium?

We have included additional information on the model protocol, specifically;

'All simulations were run at a spatial resolution of $1.9^\circ \times 2.5^\circ$, using the Qian et al., {Qian:2006wd} dataset for atmospheric forcing. The models were spun up for 1500 years to preindustrial equilibrium following an improved spinup approach (Koven et al., 2013) that allows the models to reach equilibrium after 1000 years. Simulations were then run from 1850 to 1979 under contemporary climate forcing before the onset of perturbation conditions over the following 21 years (from 1980 to 2000). Model vegetation was specified according to the MODIS vegetation continuous fields (Oleson et al., 2013).'

Page 12382, line 7: How does changing the atmospheric forcing violate the energy budget? Can't the temperature in the input file be increased by 1C? Understanding this better may help other models simulate warming experiments.

CLM4.5 calculates the surface energy budget explicitly, such that the soil thermal dynamics are driven by residual energy flux from the net radiation, latent heat flux, and sensible heat flux. We first attempted to increase the input temperature by 1C, but the simulation failed to produce soil warming comparable to the available observations. A

BGD

11, C6518–C6526, 2014

second approach we tried was to directly warm the soil by 1 C, but this approach creates unrealistic responses associated with the imposed energy imbalance. We subsequently found modifying aerodynamic resistance produced a more realistic warming compared to other approaches, such as changing wind speed.

Page 12382, Line 13-16: Well done with accounting for the intra-annual experimental treatments.

We thank the reviewer for this comment.

Page 12382, Lines 17-27: More detail about the model simulations is necessary. Did the plant functional type used in the simulation match the plant type in the experiment?

As documented in the technique note for CLM4.5 (Oleson et al., 2013; section 21.3.3), the plant functional types are specified based on MODIS vegetation continuous fields product (Hansen et al., 2003). Therefore, although it is unlikely the model simulation could match the site data perfectly everywhere, the agreement is expected to be largely reasonable at the model's spatial resolution. We have added a sentence to the materials and methods to reflect this.

Did the duration of the simulation match the duration of the experiment? For example, if the N fertilization experiment was only 3 years was only the first 3 years of the 21-year N fertilization simulation used? If the entire 21-year simulation was used then that would explain why the N fertilization response in CLM was much higher than the observations.

To account for this problem, we grouped our observationally-inferred effect sizes by experiment duration bins, where it was practical (lines 148 – 149, 156, Figure 3, Figure S1). The majority of the experimental studies were short-term (1 – 7 years long) with fewer longer term (20 year studies), and those studies dominate the effect sizes we report. For that reason, we evaluate our effect sizes with the same temporal window since experiment inception.

BGD

11, C6518–C6526, 2014

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Page 1283, Line 25: Why were the models different? Don't they have the same bio-geophysics modules?

The models do have the same bio-geophysical formulations. However, differences result from differences in the belowground carbon and nitrogen representations and the resulting impacts on leaf phenology and gross primary production. In general, CLM-CN and CLM-CENTURY behave quite differently in their soil carbon and nutrient cycles, leading to different nitrogen regulation impacts on plant productivity. The different plant productivity subsequently leads to different leaf phenology and different surface energy budgets.

Page 12384, Line 4-5: The average warming in the ESMs was different from each other and lower than the field studies. Since the models are sensitive to warming, how would the 0.3 C difference between the models influence the results? Similarly, the CLM-CN was 0.5 C lower than the observed change in temperature. This is half of the goal temperature change (1 C). What are the implications of the temperature changes not matching?

The formulations of CLM-CN and CLM-CENTURY are linear functions of the relative soil organic matter pools, and the temperature response functions are monotonic Q10 based functions. The modeled magnitude of warming was not significantly different from the observed increased soil temperatures, therefore we do not expect the modeled functional response to change qualitatively (which is the focus in this study). In addition, given the large soil carbon stocks in cold regions and that the model simulated results are opposite to empirical data, additional warming in the model would produce even stronger contrasts between model simulations and the measurements. We have added a sentence to the Discussion section addressing this point.

Page 12386, Line 21: This sentence isn't clear. If we don't benchmark using observations then what do we use?

We wished to emphasize that observations that are emergent and relatively small re-

BGD

11, C6518–C6526, 2014

Interactive Comment

sponses compared to the component processes that affect them, and where those component processes have different environmental, antecedent, or mechanistic controls, are not good tests of model fidelity. We have clarified this point in the revised manuscript.

Page 12386, Line 23: While it is important point that NEE is potentially a small difference of two large fluxes (GPP and RE), it is also important to note that GPP and RE are modeled fluxes based on NEE.

Our original point was not to evaluate the methods used to disaggregate measured NEE into inferred GPP and RH, but rather to indicate that an emergent system response that is relatively small compared to it's component drivers (e.g., NEE) is likely not a good variable to calibrate or test a model.

Discussion in general: I recommend a better presentation of the take-home messages. I also recommend synthesizing what you learned across the N fertilization and warming experiments? Are there common lessons learned in the two experiment types? Are the lessons learned that would not be found by focusing just on N fertilization or warming experiments?

Overall, I am wondering what the priorities are for CLM development based on the results from the study.

Also, the discussion uses the term "benchmarking" but doesn't providing insights into the key metrics from the study that are benchmarks for other models to use.

What metrics do the authors think that ESMs should focus on?

We have added an additional section (section 4.3, lines 697 - 744) prior to the conclusions that identifies several metrics (e.g., nitrogen mineralization, litter decomposition) that we recommend for benchmarking. Furthermore, we have also highlighted conclusions reached from the meta-analysis that could contribute to the development of the CLM-biogeochemistry codes.

BGD

11, C6518–C6526, 2014

Section 4.1: It seems that key result from the model-data comparison is the lack of an N mineralization response in the warming studies and large responses in the CLM. Why are the differences so large? What mechanisms need to be included in CLM to capture this? Why to the N mineralization response in the meta-analysis differ from other metaanalysis (Rustad et al 2001) and studies (Melillo et al. 2011)?

I would consider leading the discussion with the N mineralization response to warming because it is a core process in the climate-carbon feedback and the most striking difference between the observations and the models.

We agree with the reviewer in this case and have rearranged the discussion section to begin with the focus on nitrogen mineralization. In this new text we have addressed all of the reviewer questions posed here.

Page 12392, line 5: Other studies have found limited nitrate leaching in the CLM-CN (see Thomas et al. 2013).

We have reworded this section to note that nitrate losses are mainly from denitrification.

Section 4.3: This section does not maintain focus on the topic of barriers to experiment based model benchmarking. We know that CLM is lacking processes to perfectly simulate the globe but why is that a barrier to benchmarking. It seems that the processes that are listed should be the focus of model development through benchmarking. Overall, it seems like an odd place to provide model caveats (lack of P cycle, poor representation of N fixation, etc). The section would be more informative for other modeling groups if it explores the positives and negatives of the meta-analysis approach for benchmarking.

We have re-focused this section into four parts that address four concerns for the model versus data comparison. We also added several sentences describing criteria to be used to ensure that the imposed perturbation in the model reasonably represents the perturbation impacts in the field sites. We have also added a further section that briefly

BGD
11, C6518–C6526, 2014

highlights the positives and negatives of this approach.

Figure 2: The current size of the figure and line thickness make the figure difficult to read.

We have improved the spacing within these figures.

Figure 3: Use either GEP or GPP. One is used in the figure and the other in the caption.

We have altered this caption to reflect the use of GPP.

Interactive comment on Biogeosciences Discuss., 11, 12375, 2014.