Response to referees of bgd-11-11217-2014
October 24, 2014

We thank both referees for their thoughtful comments. Changes in the revised text are highlighted in red.

Referee #1

0. I'm not convinced that the model that has poor performance in the control site/pre-treatment site could represent the reality of ecosystem resilience. Especially, Biome-BGC seriously underestimated carbon fluxes (e.g., NEP and NPP) and the biomass accumulation for the control site, which could induce the unexpected model response (e.g., LAI in Fig 2) to the girdling in 2008.
RESPONSE: We agree this is a concern, and flag this issue (as well as suggesting a possible reason for it) more prominently in the text. See lines 37 and 434-436.

1. The authors presented each model and its parameterization in section 2.3 and 2.4. Although the text is clear, it would be better to have a table showing the differences and similarities among the three models. A parameter table for ED could be useful too…
RESPONSE: This is a good suggestion. We have added two tables: one summarizing the differences between the three models (Table 1) and the other ED's parameters (Table 4).

2. In my opinion, it is inappropriate to compare modeling results that driven by different climate data sources. Specially, climate normals were used in ED model simulations. For example, the largest difference in NEP between the control and the treatment sites occurred in warmer 2010 (Fig 4 in Gouph et al, 2013). Losing climate variability could miss the model response to the prescribed disturbance. The "sustained" carbon uptake in the disturbance site could be induced by climate events.
RESPONSE: We agree this would be ideal, and now run two (Biome-BGC and ZELIG) of the models with identical climate. The third, ED, has substantially different input requirements and, because of time and logistical constraints, we were not able to run it using the exact same climate inputs. This issue is now discussed in the methods text; in particular, we suggest that this is not likely to explain any differences in model output, because Biome-BGC and ZELIG both treat their climate data (ensembling and randomizing, respectively) that makes them quite comparable to ED's climatology. See lines 235-250.

3. Page 11218 Lines 20-23 and Page 11220 Lines 11-24: To my knowledge, most ecosystem models, such as Biom-BGC, ecosys, ED, PNET-CN, TRIPLEX, simulated much better in carbon dynamics for mature forests. Three fourth of models don't simulate stand-replacing disturbed site in NACP site synthesis studies (e.g Schaefer, K et al., 2012)…Most ecosystem models still have troubles to simulate successional trajectories of car- bon fluxes after stand-replacing disturbances such as harvests and fires. Biome-BGC has been tested several times against carbon storage and fluxes in chronosequence

eddy covariance sites (e.g., Law et al., 2003 and Bond-Lambert etal 2007), but there is still room for model improvement.
RESPONSE: We agree.

4. If the authors could organize their text following the same order during the methods and results as possible as they can, it would make the paper easy to understand. For example, I was expecting Biome-BGC rather than ED in the first paragraph of section 2.4.
RESPONSE: We have reorganized all sections so that they follow the order (Biome-BGC, ZELIG, ED) introduced in the methods.

5. Page 11224 Line 28 to Page 11225 Line 5: Is this a data-model assimilation method? I might miss something. Please identify what the search domain is? From the simulations results, I don't see the expected model performance.
RESPONSE: This was a straightforward optimization, not a full data-model assimilation. We have clarified this in the text, and removed the reference to "search domain". We also now mention one hypothesis for Biome-BGC poor absolute performance (lines 434-436).

6. Page 11226 Lines 6-10: I assume that the harvested biomass was left on-site for decomposing. Please clarify.
RESPONSE: No, the harvested biomass 'vanishes' (from the ecosystem's point of view) in all models. This has been clarified (lines 255-256).

7. Page 11227 Lines 10-15: I'm not sure how to estimate total NPP in this study. The two previous publications (Nave et al. 2011 and Gough et al. 2013) just showed aboveground wood NPP.
RESPONSE: Both observed and modeled data here are total NPP; this has been clarified (line 132).

8. Page 11228 Lines 21-22: Could the authors please give me a clue why there is no difference in the two carbon pools (leafc_storage and leafc_transfer) before and after the treatment. This could be the reason why annual leaf productions are similar. See my comment 17.
RESPONSE: This has been removed from the text, obviated by the new FPAR discussion.

9. Page 11229 Lines 14-17: Peters et al. (2012) said "At long-return intervals (200 years), increasing harvest intensity from a selective to clear-cut resulted in 11 and 10% lower mean NPP in black spruce (Figure 4) and jack pine stands (data not shown), respectively." However, in their figure 4, selective cutting (white bars) resulted in higher mean NPP compared to clearcuts in black spruce stands. I would not recommend the reference.
RESPONSE: Yes, Peters et al. (2013) say that clear-cut NPP (black bars in their Figure 4) was lower than selective-cut NPP (white bars). But their figure matches this description, with the black bar ~10% lower than the white bar for 200 yr interval. We

think the reviewer misunderstood their meaning (?), and have left this reference in our ms.

10. Page 11230 Line 4: the authors may want to say ". . . in their assumptions, parameters, and processes, . . ."
RESPONSE: Thanks–this was a mistake, and has been corrected.

11. Page 11230 Lines 7-11: I am surprised that LAI ("more or less leaf area") doesn't have effects on GPP in the three models. I probably misunderstood. But if it is true, how the models simulate the forest carbon dynamics? Could light use efficiency be changed or just be constant in the three models? Does nitrogen availability matter in Biome-BGC? How the other two gap models do? A model comparison table might help. See my comment 1.
RESPONSE: LAI absolutely affects GPP; our point in these sentences was that the models assume a fixed LUE, which is at odds with field observations and experiments. We agree that a model comparison table should help clarify things for readers, and one is now included (new Table 1).

12. Page 11231 Line 2: I don't see the Appendix 1 in the manuscript. Does it mean Table 1?
RESPONSE: Thanks–this was a mistake, and has been corrected.

13. Page 11231 Lines 8-15: Please clarify the difference between ZELIG and ZELIG-TROP? It helps little in the discussion to compare modeling results from the same model (versions).
RESPONSE: The main difference between ZELIG and ZELIG-TROP is the forest type simulated by the model: ZELIG has been parameterized and tested for temperate forests, while ZELIG-TROP has been parameterized and tested for tropical forests (both subtropical dry forests and old growth, tropical evergreen forests). This has been clarified in lines 409-410.

14. Table 1: Please change the values of maximum tree height column or the unit in the table note. Tree height of 30 m is more reasonable.
RESPONSE: Thanks–this was a mistake, and has been corrected.

15. Table 3: The authors used the same data source for Biome-BGC and ZELIG, but in table 3 the values are different. Please clarify.
RESPONSE: Yes, because the time range used was different. In the revised ms, however, Biome-BGC and ZELIG use identical climate drivers. The former Table 3 has been removed, and instead this is summarized in the text. We also discuss, in the methods, why we believe this is not a major issue–or at least, that while not perfect, it's not likely to explain any differences in model output (lines 240-250).

16. Figure 1: Please check the figure. For example, the LAI of treatment site in 2007 don't show in Fig 1 (b). Does the treatment site have the same AGB with the control site in 2007?

RESPONSE: Yes, the control and treatment sites overlap for 2006 and 2007, which is why the latter line isn't visible. Clarified in the caption.

17. Figure 2: I found that simulated LAI by Biome-BGC didn't gradually decrease. In the model experiment, the 13-14% biomass removal annually was assumed to represent the prescribed disturbance through 2008-2010. Please explain in more details?
RESPONSE: Correct: even after removing 13-14% biomass annually (= ~40% over three years), the model was able to flush (almost) a full canopy of leaves, using the internal storage pools of the simulated stand. We have clarified this in several places throughout the text, in particular the new discussion of FPAR in the discussion (lines 384-397).

18. If the models could be tested against derived annual GPP, ER (Gough et al. 2013) and soil organic carbon, it might help find why and how NEP changes after the disturbance.
RESPONSE: This is an interesting point. On the one hand, yes, it might. But on the other, our study focused on comparing model performance against the *directly observed* measurements of AGB, LAI, NPP, and NEP specifically because we wanted to use robust, well-characterized metrics. GPP and ER are, as the reviewer notes, both derived, with many uncertainties accompanying their estimates, increasing the risk of drawing a wrong inference. On balance, we'd prefer to stick to our existing metrics, unless the editor feels strongly otherwise.

*19. Page 11228 Lines 10-14 and Figure 3: it is not clear either in the text or in the figure caption that how to evaluate the model performance. Were all predicted changes for AGB, LAI, NEP, and NPP during the period 2008-2012 used? The sampling size is 20 (= 4 variables * 5 yrs) for each model? The authors may forget the third statistics (NRMSE?).*
RESPONSE: This figure attempts to summarize model performance across all metrics into a single ("Taylor") plot. We agree that the description of this figure–how it was generated, and what it means–was poor, and have clarified this (lines 322-327). We find it a useful summary figure, however, and have left it (including the RMSE component, which is a fundamental part of a Taylor plot).

20. Figure 4: Please clarify the retranslocated N? Is it leaf N, or absorbed N by plants, or N availability in soil, or N released from dead trees? If possible, please show more related N components? Theoretically, relative N availability after abnormal tree mortality events should be enhanced, as N demand might decrease and N mineralization might be improved.
RESPONSE: This figure was obviated by the new FPAR analysis, and has been removed.

Referee #2

21. My main concern about this manuscript is that the authors treat a fractional decrease in LAI like it would give the same relative decrease in FPAR. In reality these fractions are not at all the same and they depend on the absolute value of LAI and the light-

interception model you use. If you apply the most simple model, Beer's law assuming a light extinction coefficient (k) of 0.5 (FPAR = 1 – EXP(-LAI × k)) you would get the results from the table attached (Table 1). For this reason, I think that FPAR from the different models has to be included in the results and that a large part of the discussion has to be rewritten, and this means a major revision.
RESPONSE: This is an excellent and important point, and one we had completely overlooked. Thank you. While FPAR is not a standard output for any of these models (nor was it measured in the field), all use a common Beer's law formulation to compute it, and thus we can use that as a common frame of reference. The ms now discusses FPAR extensively in the discussion (lines 384-397), and has a new figure 5 focusing on it.

22. From parts of the abstract (11218:2-5, 22-23) and the introduction (11219:4-6) you get the impression that the study is about simulating the processes of tree mortality and moderate disturbances themselves, though what you is looking at the effect and recovery after a disturbance.
RESPONSE: We have attempted to clarify this (e.g. line 24).

23. 11219:5-6. "but one complicated by" is this good English?
RESPONSE: It was correct, but the sentence has been reworded and clarified. Thanks.

24. 11220:25-28. You could add that it would be for the reason to identify knowledge gaps and processes that are missing or not properly implemented for describing these mechanisms.
RESPONSE: Added (lines 99-101).

25. 11221:12. Should it be "early 1900s" and not "early 1990s"? 11221:12. Frequently disturbed?
RESPONSE: Thanks–this was a mistake, and has been corrected.

26. 11222:3. You could briefly mention the methods used for estimating LAI and NPP as they are not so straight forward to measure and are important for the study.
RESPONSE: This description has been expanded to briefly summarize the observational methods (lines 133-136).

27. 11222:4. "Gough et al." twice.
RESPONSE: Corrected.

28. 11222:10-11. Fig 2 and the text (11227:25) say 37%?
RESPONSE: Corrected.

29. 11222:21. Should it be "We tested three", or did you test several models of which you only chose to show the result from three?
RESPONSE: Clarified.

30. 11223:21-25. It might read better to divide it into two sentences.

RESPONSE: Good suggestion; changed.

31. 11224:20-21. I think you can reduce the number of "soil" in this sentence.
RESPONSE: Agreed; changed.

32. 11224:25-26. How can it be that the results were so bad (11227) if the model was optimized? Why are only one parameter (maximum stomatal conductance) significantly changed while there is no or minor changes to the rest? From where have the rest of the parameters been taken? How important are these huge deviations in absolute values for the interpretation of the results?
RESPONSE: We agree this is a concern, and flag this issue (as well as suggesting a possible reason for it) more prominently in the text. See lines 37 and 434-436.

33. 11225:14. "C pools and NPP noted above"?
RESPONSE: Clarified.

34. 11226:10. Can you be a little bit more specific about which parts of the trees that were removed from the forest and which parts that entered the litter pools?
RESPONSE: Clarified.

35. 11226:16. Is "and be subject to less competition" better?
RESPONSE: Clarified.

36. 11227. A table that summarizes these results is needed. It should include absolute values before and after the treatment and the relative change. It should also include the fraction of absorbed light.
RESPONSE: We have chosen instead to keep the absolute values summarized in the text, but include a new figure and extensive discussion (lines 384-397) about FPAR (which is not always a standard model output). We believe this addresses the substance of this comment, albeit slightly differently than requested.

37. 11230:4. Repeated word "processes".
RESPONSE: Thanks–this was a mistake, and has been corrected.

38. As said above, a lot of the discussion has to be revised as you assume that the relative change in LAI can be interpreted as change in FPAR. E.g. the conclusion "It was instead Biome-BGC that best maintained light absorption" (11230:21-22) is wrong and you have not shown that.
RESPONSE: Yes, the discussion has been extensively revised to reflect this insight (lines 384-397 and elsewhere).

39. 11232:14. Take away "remains".
RESPONSE: Corrected.

40. 11232:17-20. I would not keep the philosophical ending of the discussion.
RESPONSE: That was a mistake—our apologies! Removed.

41. Another aspect that could have been discussed is how the experiment is different from a real disturbance and what impact that has on the results and if the models might do better than the experiment in some respects?
RESPONSE: This is an interesting point, and was addressed by Gough et al. (2013, cited in references). We don't think it needs significant discussion here, but are happy to do so if the editor thinks it necessary.

42. There are a lot of question marks in the conclusions!
RESPONSE: Yes-this study raises more questions than it answers, we agree. We have simplified them a bit (line 454).

43. Table 1:key. Those HT max values seem high, should the unit be cm?
RESPONSE: Corrected.

44. Figure 2. "Model performance in replicating results", I think that the English could be tightened up here. I think that it is better to explain what is shown directly than refer to Fig 1. If you include a table with absolute values you can refer to it. It would be good to add FPAR.
RESPONSE: The caption text has been tightened and clarified. Re FPAR, it is not included since it's not a direct model output, but is shown (see response to comment #21 above) in the new Figure 5.

45. Figure 3. Is this the most straight-forward way to show these results? I am not a statistician and for me it is a little bit more complicated than necessary.
RESPONSE: This figure attempts to summarize model performance across all metrics into a single ("Taylor") plot. We agree that the description of this figure–how it was generated, and what it means–was poor, and have clarified this (lines 322-327). We find it a useful summary figure, however, and have left it (including the RMSE component, which is a fundamental part of a Taylor plot).

46. Figure 4. Better to have the units directly to the y-axis. "Shading shows meteorological variability", how can there be a meteorological variability in N and C stocks? "vertical dashed line", it looks solid grey to me.
RESPONSE: This figure was obviated by the new FPAR analysis, and has been removed.