General Comments

In the manuscript "Modelling decadal trends and the impact of extreme events on carbon fluxes in a deciduous temperate forest using the QUINCY model", Thum et al. explore how using in-situ measurements of vegetation traits to parametrize the nitrogen cycle-enabled QUINCY model improves flux simulations at the Borden Forest flux tower. The QUINCY model is modified to allow for a delay in leaf chlorophyll development and the model runs produced in nine different simulations of varying parametrization and nutrient cycle implementations.

When using locally calibrated parameters and enabling the nitrogen cycle in addition to the carbon cycle, QUINCY performed better against observations, especially for GPP, but still lacked some key behaviors – namely it failed to capture drought response and legacy.

Overall, the manuscript is of a high quality, presenting a well-developed study in a clear and scientific manner. It is a substantial contribution to the literature. I recommend publication following minor revisions.

Specific Comments

1. I like the specific statement of the research questions in the study at the end of the Introduction. However, these are not referred to again. I would suggest a section of the Discussion is reformulated to explicitly restate these questions and then explore the evidence found for each one. This would provide a good narrative throughout the manuscript and help in synthesizing the findings of the study.

2. Among the limitations of the study is the level of representativeness of the chosen PFT for the flux tower site. According to my reading of Thum et al. (2019) and its supplementary material, QUINCY has the capacity to model an individual gridcell as nested tiles of different PFTs. Does this not provide the necessary flexibility for QUINCY to represent a more accurate mix of vegetation when modelling Borden Forest? What was the inter-species variability in the leaf-level parameters and how much did the species-weighted average differ from the species level values?

3. Studies have shown that a simple representation of carbohydrate pools can help to model drought legacies (Guo et al., 2020; Jones et al., 2020). QUINCY has a representation of these NSC pools and so might be expected to capture the drought legacy in 2008. As such, I'd like to see Section 4.4 in the discussion expanded to discuss in more detail, or at least more focused on, the reasons why QUINCY is

unable to simulate drought and drought legacy. Exploring the processes behind the observations is interesting but this should be framed from the perspective of model evaluation and used to synthesize our understanding of QUINCY's performance.

4. Table S1 is referred to throughout the manuscript at greater frequency (12 times by my count) than every figure and table in the manuscript except Figure 4 which is also referenced 12 times. Alone, it is referred to more times than all three of the tables included in the manuscript! I would suggest moving this table to the manuscript or otherwise incorporating this information in a manner that does not result in the jarring requirement of frequently referring to the supplement.

5. Figure 2 shows the mean yearly values for NEE, TER, soil temperature and soil moisture. This is discussed in detail. My feeling, considering the soil temperature plot and the location of Borden Forest, is that snow cover may be playing a substantial role at this site over winter. Note that the observed soil temperature maintains a constant temperature through winter, consistent with an insulating snow cover. This may also explain the model's earlier TER response as QUINCY fails to simulate any snowmelt period suppressing soil respiration despite increasing temperatures (Teklemariam et al., 2009). Was this potential role of snow considered?

Technical Comments

1. L20: Delete "the" from before "ecosystems".
2. L27: Longer growing seasons compared to what?
3. L30: Delete "the" from before "vegetation functioning".
4. L33: Delete "the" from before "forests". Note this is a recurring issue as per comments 1 and 3. Since we are not discussing specific forests or ecosystems, there should not be a definite article "the" in these instances.
5. L51: Add "an" before "increase" and a "the" before "land carbon sink". Modify "changing of" to "change in".
6. L61: GPP has already been defined.
7. L69: Delete the comma after "Borden Forest" or add a comma after "continuous data".
8. L78-79: This sentence needs to be reformulated to make sense.
9. L81: Chl_{Leaf} has already been defined.
10. L83 and throughout: Make sure references are formatted correctly, namely the brackets.

11. L125: GPP has already been defined.

12. L163: The acronym QUINCY has already been defined.

13. L202: Define or explain "OCN".

14. L228: Delete "an" from before "the slope".

15. L243: Add "and" between "humidity" and "wind".

16. L247: Was a single random year used repeatedly in the spin-up or was it a 500 year time series constructed by randomly selecting a year of [CO2] and meteorology for each of the 500 years?

17. L249: What data was used for the meteorology between 1901 and 1996 in the transient runs? ERA-5 is mentioned earlier but this dataset begins in 1940.

18. L264: Change "showed" to "shown".

19. The model abbreviations are quite long and complicated – is there a way to condense them while maintaining the information?

20. L272: As I understand it, LAI was a parameter that was calibrated and this can be seen by the improved seasonality of LAI in Fig 1c&d. Was there no capacity to correct the magnitude of LAI which is too high in summer and too low in winter?

21. L286: Figure S2 shows 2013, not 2014.

22. L297: "The observations show more shallow decrease" - I think this should be "The observations show a more gradual decline in increasing GPP before the peak" or similar?

23. L304: "parameterizations is in" should be "parametrizations are in".

24. L305: It may be worth clarifying that this refers to different simulations within the "C-only" model framework only, not between the "C-only" and "C-only,fix" model simulations.

25. L309: Perhaps specify that the "more accurate representation" is better timing of senescence?

26. Figure 1: GPP and LAI have already been defined, and "leaf chlorophyll" has not been defined as "Chl" but as "Chl$_{Leaf}$". I also do not think that the colors of each model simulation need to be spelled out in the caption as the legend already provides this information. It should be "... leaf chlorophyll WHICH has been smoothed...".

27. L313: Perhaps a column should be added to Table S1 containing the percentage figures of under/overestimation.

28. L316: Should it be "with only a 1.4% larger value"?

29. L320: I think this should refer to Table S1, not Table S2.

30. L322: TER has already been defined.

31. L330: "Table 1" should be "Table S1".

32. Figures S3, S4, and S5 are often used to illustrate points that are difficult to parse from the plots (namely referring to fluxes being over/underestimated in certain months). The plots are too noisy for me to clearly and easily identify these statements so I suggest a different method of plotting this data to illustrate the points made.

33. L339: "The early season pattern observed in the simulated NEE is attributed to heterotrophic respiration". Is not the overestimation of early season NEE due to a GPP that is too low from DOY ~75, as well as TER being too high between DOY 50 and 100? This is what I read from Figure S6.

34. L348: Missing bracket after "(Fig. 2."

35. L355: This sentence is not properly formatted.

36. L365 and surrounding paragraph: This paragraph could be condensed. It is also unnecessary to specify (C-only:LAI&Chl) and (CN:LAI&Chl) after each reference to the models as it has already been stated that these are the models referred to from Line 318 onwards.

37. L376: This statement about SLA is disjointed from the rest of the text.

38. L379: TER is discussed first in Figure 4 but in the plot, GPP comes first. I would ensure these align to improve readability.

39. Figure 4: In this instance, I think it might be better to have the observed and simulated on the same plots, with the different facets instead for the years – as it stands, the figure is better as a comparison between years than for assessing model performance, which I believe to be its main message.

40. Figure 4 and Figure S9: As in other plots, could the standard deviation be plotted as a shaded area around the observation means?

41. L392: Move the sentences about GPP to the next paragraph which discusses GPP more widely, rather than in this paragraph which is about TER. Section 3.4 in general could use some work to improve the flow of the section and maintain a narrative throughout instead of the current text which tends to jump between the discussed variables with little reason. Either discuss each variable in turn, or discuss each time period in turn.

42. Table 4: The QUINCY LOS (GPP) is incorrect – it should be 146 days if calculated as the difference between the mean SOS and mean EOS as the other table elements seem to imply (or perhaps this is coincidence and it is actually the mean length of each individual growing season, in which case this should be clarified).

43. L417: "LAI based estimates" is misspelled.

44. L420: "takes place in average" should be "takes place on average".

45. L429: Clarify what is meant by "make use of different spring and autumn periods".

46. L438: There appears to be a significant breakpoint in GPP around 2009. Are there any explanations for why this might be? Is this a recovery from the 2007 drought? Is there any potential reason why every year post 2010 has higher GPP than any year before? Why were the final 5 years removed as a test?

47. L445: Why is LAI not plotted in Figure S12 if it is discussed here?

48. L451: Can the differences in IAV be quantified somehow, for instance comparing the standard deviations?

49. L469: Incorrect parentheses for reference.

50. Section 4.1: I'd like to see more discussion here about how the continuous LAI measurements improved the model performance. What has been learnt from this study and how can these lessons be implemented in the model? Does the senescence parameter need to be modified in the standard implementation of the model? Do we need to test the model tuned with continuous LAI values at other sites?

51. Section 4.2: Again, I think there could be more discussion here about what the simulated Chl_{Leaf} values can teach us regarding QUINCY and future improvements. What does it mean for the model simulations that Chl_{Leaf} peaks early? What might this imply for QUINCY applied at other sites or globally? While not the objective in this study, what might be learnt if we did attempt to capture the timing of the maximum leaf chlorophyll? What would the tradeoffs be?

52. L511: "observed transitions is more smooth" should be "observed transitions are more smooth".

53. L514: I'd change "the model used here" to "QUINCY".

54. L515: Missing word between "Testing model performance" and "a TBM designed".

55. L570: "It occurs most pronounced" is not grammatically correct – replace with "It occurs most prominently" or "It is most pronounced".

56. L571: "Based on the data available then…" reads as if referring to the 2008 drought due to the prior sentence. I'd recommend "Based on the data available to Lee et al. (1999), they found …" or similar.

57. L583: Delete comma after "exudates".

58. L587: Typo in "whereas is was".

59. L598: "QUINCY does simulate" should be "QUINCY does not simulate".

60. L600: There are quite a few statements throughout the manuscript similar to the sentence here: "One additional cause of model failure might be that the canopy light-saturation point does not reflect the observations, however, there is not robust evidence that this is the case." These require at least some explanation of what potential sources of evidence were explored and discounted.

61. L605: Add "an" before "impact on the N cycle".

62. L619: Add "in" after "cause a change".

63. L623: Add "an" before "unrealistically low value".

64. L630: Add "the" before "N saturated case".

65. L631: Delete "and" before "in line with".

66. L633: Delete "of" in Section 4.8 header.

67. L642: Add "A" before "rain gauge".

68. L648: Specify that the "long time spans" is referring to long time series of observations.

69. L648: Specify that the "use of leaf chlorophyll content and LAI" is in parametrizing the model.

70. L658: Change to "attributed to an increase in PAR which is not visible in the shortwave radiation forcing for QUINCY".

71. L661: Delete "is" from before "paves the way".

References

Guo, J. S., Gear, L., Hultine, K. R., Koch, G. W., & Ogle, K. (2020). Non-structural carbohydrate dynamics associated with antecedent stem water potential and air temperature in a dominant desert shrub. *Plant, Cell & Environment*, *43*(6), 1467–1483. https://doi.org/10.1111/pce.13749

Jones, S., Rowland, L., Cox, P., Hemming, D., Wiltshire, A., Williams, K., Parazoo, N. C., Liu, J., da Costa, A. C. L., Meir, P., Mencuccini, M., & Harper, A. B. (2020). The impact of a simple representation of non-structural carbohydrates on the simulated response of tropical forests to drought. *Biogeosciences*, *17*(13), 3589–3612. https://doi.org/10.5194/bg-17-3589-2020

Teklemariam, T., Staebler, R. M., & Barr, A. G. (2009). Eight years of carbon dioxide exchange above a mixed forest at Borden, Ontario. *Agricultural and Forest Meteorology*, *149*(11), 2040–2053. https://doi.org/10.1016/j.agrformet.2009.07.011