General comments

The study of Thum et al. investigates carbon and nitrogen interactions in the Ontario's Borden Forest Research Station, using in-situ measurements to parameterize the QUINCY model. It evaluates carbon flux simulations over 22 years, finding good alignment in some metrics like GPP but identifying key discrepancies in ecosystem respiration trends and legacy drought impacts, underscoring the need to improve TBMs.

The manuscript addresses an important topic: the representation of carbon and nitrogen interactions in TBM models. Overall, it is well-written, and the figures—despite a few editorial issues—are clear and effectively support the results presented. However, I found the study's objectives difficult to discern from the Abstract. The Introduction also requires substantial revision, as it sometimes lacks logical flow, and the paragraphs don't fully cohere. For instance, the first paragraph focuses on the importance of nitrogen, followed abruptly by a discussion on changes in the growing season due to warming, and then by a mention of the value of long-term observations for capturing anomalies (anomalies of what?). The Introduction feels like a series of loosely connected points without a clear narrative thread, which makes it challenging to understand the paper's aim until the objectives are listed in the final paragraph.

Regarding the content of the paper, there are two key aspects that should be addressed:

1. **Quality of Eddy Covariance Flux Post-Processing:**
   The study relies heavily on estimates of GPP and TER derived from eddy covariance measurements, which are not directly measured values. It is essential to assess the quality of the partitioning and gap-filling methods used. A study from 2004 is referenced for both flux partitioning and gap filling, yet there is no description of this approach, its advantages, or why it was chosen. Given that GPP estimates are highly sensitive to partitioning methods, it's critical to first establish a solid foundation, showing that the best possible gap-filled and partitioned fluxes were derived before making further interpretations about model structure or other underlying factors.
2. **Snow Cover Effects:**
   Considering the site's geographic location, persistent winter snow cover is likely; if this is not the case, it should be explicitly mentioned. However, the manuscript does not address snow cover or its potential impact on carbon flux processes. This is especially relevant in the shoulder seasons, where the authors discuss discrepancies between modeled and observed fluxes. Factors such as snow effects, soil freezing and thawing, and changes in soil-air temperature decoupling due to snowmelt could significantly influence these processes but are omitted in the analysis.

Below, I provide more specific comments on the manuscript.

Specific comments

Line 10: Please also report the RMSE (and not just r2) when reporting the model performance.

Line 10: You mentioned how you parameterized the model but not how the model was improved.

Line 11: Would be interesting to know the magnitude of this increase

Line 11: NEE not defined yet

Line 35: Grammar (“are” missing)

Line 42: Would be important to add here why the current representation of N limitation of photosynthesis is not sufficient (since this is such a key aspect in this paper)

Line 43: For example, which responses?

These aspects are important to be clarified to convince the readers why “it is paramount that the effects of N constraints on plant productivity are accurately simulated”.

Line 50 and 52: Do these observations overlap? Both in terms of time and space? How is the increase in LAI and the decline in N related?

Line 79: Grammar

Section 2.1: Please add a description of the understory vegetation (species and cover). This information is relevant to your discussions of model limitations that we read later in the paper.

Lines 101-104: Is this the species composition within the flux footprint?

Line 107: Mean over what period?

Line 118: Used for what? For calibration of the model? For validation?

Line 126: A brief description of the gapfilling and partitioning method should be given here and the justification why such method is selected over other existing methods.

According to Barr et al 2004 “seasonal onset, rise and fall of photosynthesis from the *F*NEP time series based on the parameter *Px* in the rectangular hyperbolic model “was this the case too in this study?

Line 128: Grammar

Line 133: How exactly was this scaling done?

How well do the ERA5-Land precipitation product and measured precipitation at the site compare? Perhaps a comparison can be added to the supplementary.

Line 193: Typo

Line 274: Based on what was this level of T selected for the adjustments?

There is a typo just before the caption of Table 2.

Figure S2 has a typo in the legend of panel b (should be LAI not GPP)

In Figure S2 the panels already show GPP or LAI. I would not repeat this again in the legend. Also, the dashed lines can be removed from the legend (to make it less crowded) and keep their description in the caption.

Figure S3- 5 the axis label should be “modelled” to be consistent with the x-axis which says “Observations”. The figure title is already mentioning which model was used, not needed to mention this in the title and in the axis label.

Figure S3- 5 wouldn’t it make more sense to display mean daily temperature as the third value rather than the number of the month? What is the reasoning to use month here? Maybe can briefly be added to the caption.

Figure S3- 5 it is hard to judge quantitatively the comparison of different model performances. Maybe at least the r2 values can be displayed in the panels?

Figure S8 very hard to read the figure. Consider increasing the font please.

Line 339-345: Could this inaccuracy in modelling the soil temperature be because of the snow? What is the contribution of snow cover at this site? We see from colder sites that snow has an insulating effect on the soil that decouples its temperature from air temperature. If direct measurements are not available at the site perhaps you could explore available remote sensing products (e.g., MODIS/Terra (MOD10A2) and MODIS/AQUA (MYD10A2) (Hall and Riggs 2021) Snow Cover 8-Day L3 Global 500m SIN Grid, Version 6 dataset, which provides maximum snow cover extent at 8-day temporal resolution and 500m spatial resolution).

Line 460-464: Here the Results are repeated. Instead, there should be a discussion of what underpins the observation that although modelled LAI is overestimated, modelled GPP in summer is underestimated.

Section 4.2 I suggest dividing parameters by structural from photosynthetic traits.

Line 491-493: What explains it if the SLA was overestimated but GPP underestimated?

Line 500: “instead, we are estimating the tree traits per average individual for a deciduous forest.” Not clear to me what this statement means. Because the methods section (2.2.3) only mentions “we used a species-weighted canopy average of the leaf-level parameters, based on the species composition of the forest „ and is not clear how the parameters are weighted and then aggregated (?). Adding a mathematical description here would be very helpful.

Line 510: Which drought occurrences?

Line 515-517: Grammar check and re-writing needed. Sentence is not clear.

Line 525: “The tree species composition has undergone changes at the site during our study period, e.g., the red maple was reported to have coverage of 36 % in 1995 (Lee et al., 1999) and 52 % in 2006 (Teklemariam et al., 2009). The impacts that these changes in the tree

composition have on the carbon fluxes could be studied by a demographic model with sufficient granularity in the description of tree functional diversity“

Since species compositional shift was not really addressed in this study would suggest to remove this context from the Introduction as currently it reads as if this is one of the aspects this paper addresses.

Section 4.4 it is not clear if the model fails to simulate the (legacy) effect of drought because of its structure (representation of the carbohydrate pools) or the lack of precise soil moistures estimations which directly affect modelled CO2 fluxes. Could this not have been specifically tested (for this particular case of drought conditions) by first calibrating the model using observed soil moisture measurements and testing whether GPP simulations improved during drought? If the model´s limitations during drought is a focus of this paper it deserves a more systematic approach to test this.

Line 556: What is meant here? Isn't a depth-resolved soil texture provided to the model? Otherwise, what description of the soil physics exactly is lacking here?

Line 570-576: Yes, this hysteresis in TER response to temperature (early season lower respiration than late season) could reflect the seasonal patterns of photosynthate allocation to roots. Tree girdling studies have shown that seasonal pattern of below-ground C allocation may be more important than soil temperature in determining root respiration (see for example Högberg et al. 2003) and that earlier in the season respiration from the soil is mostly due to heterotrophic respiration.

Line 591: Increase over time? Sentence reads incomplete.

Line 597: So the PAR that was used as an input to the model was not correct? The model would clearly use PAR, so why give it the wrong forcing?

Line 603: You mean at this site? It is not clear how these findings are relevant to this study.

Line 600-607: The discussion on the ozone effect comes out of blue here, unless my previous comment is clarified.

Lien 614: What is meant by “differences between the annual GPP and carbon balance“?

Line 619 and 620: Grammar

Section 4.8: I would not finish the manuscript on listing technical shortcomings. What have we learnt from long term ecological response of such an ecosystem, observations and model results combined? And what is the outlook under further changes in the climate? (e.g., predicted potential increase in temperature and dryness).

The Conclusion section can be shortened.

References

Hall, D. K. and G. A. Riggs. MODIS/Terra Snow Cover 8-Day L3 Global 500m SIN Grid, Version 6. Distributed by NASA National Snow and Ice Data Center Distributed Active Archive Center (2021).

Högberg, P., Nordgren, A., Buchmann, N. et al. Large-scale forest girdling shows that current photosynthesis drives soil respiration. Nature 411, 789–792 (2001). https://doi.org/10.1038/35081058