1 Reviewer 1

GENERAL COMMENTS:

[...] The authors find strong correlations between modeled and observed productivity data and recreate fertilization and nitrogen dynamics found in other studies, demonstrating their success in adding shrub PFTs to the QUINCY model as a basis to address further questions, though there is less evidence that the model is directly improved by the new PFTs versus the existing ones. As there is no new measured data and the manuscript hinges on the QUINCY model results, without a convincing improvement in the model, the scientific content is limited to measuring shrub-specific carbon and nitrogen cycling. The title of the work suggests the authors also focus on those sections, but there are several statements that would be overextended based on the actual content of the paper without demonstrated model improvements (e.g. lines 16-17 "underlines the importance of... shrub PFTs in global land surface models..."). While this work could be an important step to improving a strong model linking Arctic shrub processes and climate dynamics and the science has merit, one of the major arguments made is unclear, weakening the impact. If the explanation of the impact on the model and the clarity of arguments were to be improved, the manuscript would be a valuable contribution to Arctic vegetation-climate modeling and a foundation to include more detailed permafrost dynamics to vegetation models. Any issues with the paper are not in the science but in how it is conveyed and all the most significant issues occur in trying to make general statements about the model. As written, the major scientific contributions of the study are the creation of a tool to evaluate shrub-specific climate-vegetation interactions in the Arctic and using that tool to specifically analyse how shrub growth and carbon cycling are impacted by nitrogen availability.

We gratefully acknowledge the supportive yet critical review and agree to clarify the study findings. Please find our detailed responses below (in black).

The paper's focus is not on the explicit model validation, but rather on the impact of shrub vegetation in the Arctic. Thus, respective statements are overextended and we reformulate. For example, we propose the following for line 18/19 ('These finding underlines the importance of including evergreen and deciduous shrub PFTs in global land surface models to accurately predict ongoing changes in the Arctic C cycle'):

Overall, the study shows the impact of evergreen and deciduous shrub vegetation on the Arctic carbon cycle and ongoing changes, recommending the inclusion of shrub plant functional types in global Earth system models.

SPECIFIC COMMENTS:

In terms of specific issues, one of the primary examples of lack of clarity is the improvement of the model by including shrubs versus using QUINCY's inbuilt PFTs. While the title of the paper focuses on the specific carbon and nitrogen impacts (questions 2 and 3), about half the paper is devoted to the implementation of the new shrub PFTs and the abstract ends with a claim that introducing shrub PFTs improves the accuracy of models. Given that, when compared, the median GPP values and total carbon between shrubs and trees are not significantly different, it would be worth emphasizing other information to show an improvement in model performance (e.g. Is the R2 improved? Are there verifiable site-level differences?, etc.). Adding a line for the model using the tree PFT to Figure 1, for example, would add weight to the significance of the paper if the R2 is improved. Since the skewness is different between shrubs and trees, does the shrub skewness better match the observed values than the tree skewness? While it is important to include the similarities in the medians, there are other statistics that can be compared to highlight what does change. Or in the other case, if measures of model accuracy are included purely for validation and do not make any statements about the general QUINCY model, it would be worth being more specific about that; even just changing a section title could improve clarity (e.g. 3.1 Validation of modeled vegetation productivity). One of the main strengths of the paper is to serve as a foundation to expand the work and extend the new model, and if that strength were explicitly conveyed, it would focus the reader on what was improved rather than the overall performance of the model.

We understand your concerns and try to clarify and improve the mentioned aspects. Section 3.1 is intended to showcase the shrub productivity at the study sites and we do not aim to validate the model performance overall. However, you are right, by adding the (default) tree GPP in Fig. 1 (see below), we can show the value of including the NE shrub PFT. We will adapt the text in Sec. 3.1 and the abstract accordingly. We will change line 163-167 to:

The modeled shrubs capture a reasonable monthly mean seasonal cycle of GPP at half of the validated sites (six out of twelve

with a 50 % measured seasonal cycle), when compared to the ABCflux site measurements in Figure 1. Modeling with the new NE shrub PFT improves the mean R2 value to 0.50, compared to the default (tree) PFT across nine sites (dashed line in Fig. 1 and Table A2). In contrast, simulations with the BD shrub (7 sites) mostly underestimate the measured seasonal cycle of GPP, showing no improvement compared to the performance of the default (tree) PFT (mean R2=0.6).

and line 172/173: *Here, the MODIS GPP products show a higher correlation towards the model with a mean R2 value of 0.71 at NE shrubs and 0.50 at BD shrubs (Table A3) than compared to the site measurements. Generally, this is an improvement for NE shrubs compared to the NE default (tree) PFT (R2=0.60).*

We will shift the discussion on the CN vs. C-only model configuration to Section 3.3 as suggested.

We will adjust line 192/193 to:
The annual GPP at shrubs has a much wider inter-quantile range than the distribution at tress and graminoids i.e. a higher variability. The distribution is more symmetric (skewness=0.6 vs. 1.2 (trees) and 1.0 (graminoids)) having shorter tails and a flatter peak (kurtosis=-1.3 vs 0.3 and 0.0).

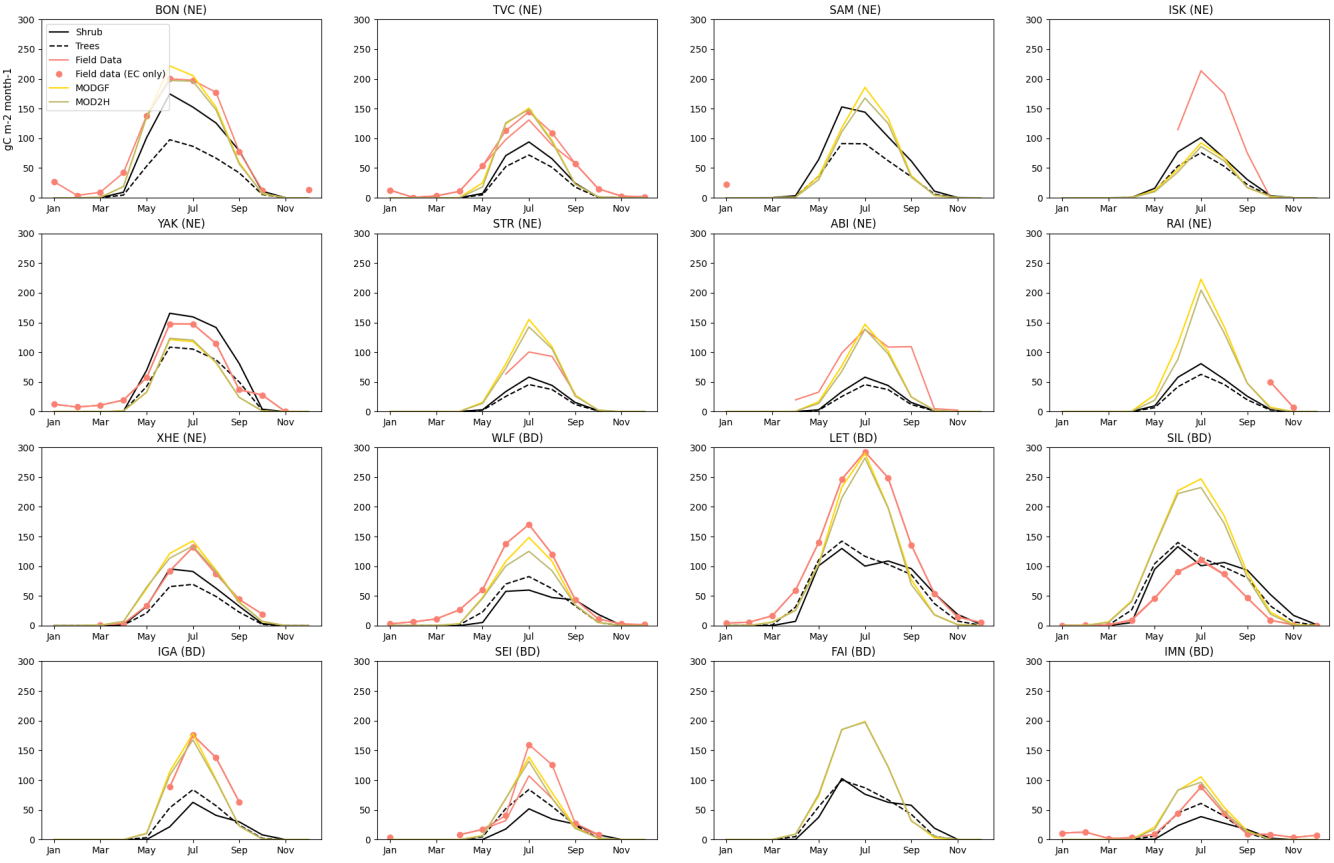

Figure 1. Multi-year mean seasonal cycle of GPP [gC m-2 month-1] by measurements (field data and MODIS satellite data: MOD17A2HGF and MODIS17A2H) and model with new shrub PFT, tree PFT and graminoid PFT at 16 sites (BON: Bonanza (US), TVC: Trail Valley Creek (CA),SAM: Sammaltunturi fell (FI),ISK: Iskoras (NO), YAK: Yakutsk (RU), STR: Stordalen, ABI: Abisko (SE), RAI: Raisduoddar (NO), XHE: Neon Healy (US), WLF: Wolf Creek (CA), LET: Lettuoso, SIL: Siikaneva (SE), IGA: Igarka, SEI: Seida (RU), FAI: Fairbanks, IMN: Imnavait (US)) at individual time periods where field data is available.

Another key point that could be made clearer is the statement in lines 8-9, 165, and 312-313 that model nitrogen dynamics may be overly restrictive. This is a major point in the paper that relates to stronger/weaker correlations between modeled and observed data, even being referenced in the abstract, but the logic in reaching this conclusion is somewhat murky. This is not the only possible explanation for why the C-only model had higher R2 values than the CN model in half the sites, and while line 313 suggests issues in modeled nitrogen fixation, any discussion thereof is in the section on CO2 fertilization and is decoupled from the statement. For the actual logic used to reach this conclusion, to the best of my understanding, in addition to the biological nitrogen fixation mention the paragraphs about permafrost dynamics following lines 312-313 are meant to serve as an example: the reduced nitrogen limitation in the permafrost model had a stronger correlation with observed data. If this is true, the link between the two pieces of information is broken by how the next paragraphs are structured, discussing model parameters before any mention of model improvement. It is also handled far too late in the manuscript given where the claim is first introduced. Because the paper uses a "Results and Discussion" section rather than breaking them up, data and interpretations are interspersed. This means any claims need to be explained when introduced and any callbacks to other statements need to be explicit.

We agree that this aspect on the nitrogen dynamics is overextended. The C-only model clearly shows better comparison to-

wards observation data but we cannot overly prove that the nitrogen dynamics are too restrictive. Therefore, we change line 8-9 ('However, at 50% of the study sites the model underestimated observed GPP due to too strong simulated nitrogen limitation.') to:
However, local heterogeneous ecosystems can not be captured by the model hydrology or represented by the shrub plant functional type yet.

Line 165 is changed due to the re-writing of section 3.1 (see comment above). We shift the discussion of the C-only model to Section 3.3 showing a Figure with the measured seasonal cycle of GPP compared to the CN, C-only and the permafrost model version GPP there (modifying Figure S17). Line 312/313 ('However, our results suggest that the N dynamics in the model may be overly restrictive, and could be improved by refining key processes such as BNF to better reflect Arctic conditions') can be supported by additional measurement data of the C/N ratio from the TRY database (Kattge et al. 2020). We extracted 89 measurements from Betula nana and 36 measurements from Betula glandulosa in the Arctic. The C/N ratio of 24.0 (median) estimated by the C-only configuration at 7 stations (with deciduous species) agrees much better with the measured values (26.06±2.0, 22.56±2.1). The CN model median value of 29.06 instead overestimates the measured C/N ratio, i.e. underestimates the leaf N concentration supporting our hypothesis.

A technical issue with clarity is the way the supplemental material is and isn't referenced. There are 18 supplemental figures (compared to 6 in the main body and 2 in the appendix), some of which have more bearing to the actual contents of the paper. The most problematic instance is in lines 204-205, in which the reader is instructed to see Supplemental Figure 3 for R2 values that are referenced later in the paragraph (line 209: "might explain the low correlation") but are never summarized or listed in the main text, which could arguably be seen as a violation of the policy that supplementary materials cannot present findings or interpretations beyond the contents of the manuscript. As one of the review criteria is whether the quantity and quality of supplementary materials are appropriate, I would recommend reviewing the figures and determining if any of them are unnecessary (e.g. is the ubiquitous atmospheric CO2 graph in S12 needed?), are needed to be summarized or included in the manuscript (e.g. S3 R2 values), or need more information in the legend or caption (e.g. grey and black lines in S14), particularly since the code and data are available.

We believe the supplement material provides important additional details on the model findings. We will improve the structure of the figures to remove sub panels where some are not referenced in the main text and reduce the number of SI figures.
The R2 values in Figure S3 are summarized in line 203/204 of the main text ('For all shrubs, the change of NH4+ uptake explained 50-70% of the GPP change compared to trees (see R2 in Fig. S3)').
We agree that the CO2 graph in S12 is ubiquitous and can be removed.
Table S1 can be removed as well because it is not referenced.
We apologize for two missing legend entries in S14, these will be added.
A modified version of Figure 17 will be re-included in the main text.

TECHNICAL COMMENTS:
Other, smaller issues are primarily concerned with readability. There is a missing citation in Table 2 and the y-axis on Figure 5a is unlabeled but could represent portion of the month included in the growing season, biological activity in the month, portion of years in which the growing season extended into the month, or something else, changing the meaning of the figure. There is a lack of consistency in patterns between panels and between figures/tables. In Figures 2 & 3 (as well as some supplemental figures), the color representing shrubs and graminoid values swaps between panels a and b. The order in which sites are listed changes in every table and most figures. In both cases, maintaining consistency would increase the readability of the paper. In addition, Figure A1 provides important enough context that it may warrant inclusion in the main body of the manuscript, but if not, some explanation of the site codes needs to be present outside of the appendix. Lines 205-206 compare evergreen shrubs to shrubs, which is confusing - the authors may mean trees (based on context) or may mean deciduous shrubs - either way it should be clarified. The values in the paragraph in lines 219-222 may be swapped as the aboveground total is approximately twice the presented overall biomass. The NaN in Neon Healy vegetation type in Table A2 merits an explanation in the caption (similar to the Table A3 caption). The paragraph at lines 113-124 repeatedly cites the same dataset and would be better served

by a statement like "using the methods of Palmtag et al. (2022), [Values used]..." to not appear more reliant on the source than the paper actually is. The manuscript would benefit from detailed copy editing as there are several small grammatical and punctuation errors, some (e.g. a missing comma after "yet" in line 249) that impact meaning, while most (e.g. wrong "its" in line 5 or "foilage" instead of "foliage" in line 272) do not.

We will check and improve the citations, grammar and punctuation.
We apologize for missing consistency among the figure colors and the site lists. We will harmonize that.
Also, we agree to shift Figure A1 to the main text.
We will harmonize the order of the site lists and used color bars.
We will correct the sentence meaning in Lines 205-206. Indeed, it compares evergreen shrubs with the default (tree) PFT.
We will improve the readability in lines 113-124 like suggested.
Regarding table A3, NaN values mark data with too small data size as it is mentioned in the table caption.