Reply to Reviewer comment 2

General comments

...

This study and model development are exciting and have potential for an important contribution, though changes such as clarification and better justification of the conclusions and improvements to readability are needed prior to publication. In particular, the manuscript would benefit from some changes to address lack of clarity and better support the main conclusions being made. These include 1) more explicitly demonstrating the added value of the new shrub PFTs relative to existing tree parameterizations to show how the model is improved by the new shrub paramterizations; and 2) evaluating physical processes in the model such as snow depth, soil temperature or moisture, and active layer dynamics against observations from the sites to better support the claim that nitrogen limitation is the main source of modeled GPP bias rather than potential model inaccuracies in snow and soil conditions. The authors mention that shrubs differ from trees in allocation, phenology, and rooting characteristics, but it is less clear how much predictive skill is gained relative to the existing tree parameterization. A more direct comparison showing the extent to which the shrub PFTs improve model performance versus simulations using the default tree PFT would be helpful, for example by including comparison of observations with both shrub and tree PFT results (this could be added to Figure 1). Clearly demonstrating how the model is improved/changed with the addition of these two new shrub PFTs would make this a more impactful contribution.

The authors thank the reviewer for reviewing our manuscript. We greatly appreciate the time and effort invested in providing detailed and constructive feedback, which will be very helpful in improving the manuscript. Please find our detailed answers below in black (reviewer comments are in brown).

Specific Comments

1. One aspect that could use clarification is how the new shrub PFTs improve the model relative to previous tree configurations, given that this is one of the main claims of the paper. Many of the reported differences between shrub and tree simulations appear relatively modest (e.g., non-significant differences in median GPP, 16% differences in mean GPP, and small differences in soil carbon storage). If the authors can better demonstrate the extent to which the new shrub PFTs improve model predictions or alter ecosystem carbon budgets relative to the existing tree parameterization, this would help justify the conclusion that shrub representation is necessary for accurate simulation of Arctic carbon cycling (a conclusion that makes sense, it just needs better support).

We will follow the suggestion and add one line for the tree and gramonoid GPP to Figure 1 (removing the line for the C-only simulation). With calculating additional R2 values we can demonstrate that at least the modeled needle-leaved evergreen shrubs perform better than the default PFT when compared to the observation products. However, we explain the challenges in modeling the heterogeneous ecosystem exploratory at some site. For example, the Lettuoso site in Finland is covered by large forest additionally to shrubs, thus the default PFT performs better. We propose to change line 163-167 accordingly:
The modeled shrubs captures a reasonable (monthly mean) seasonal cycle of GPP at half of the validated sites (6 out of 12 with 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 9 sites (dashed line in Fig. 1, Table A2). In contrast, simulations with the BD shrub (7 sites) mostly underestimates the measured seasonal cycle of GPP which is no improvement compared to the performance of the default 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).*

2. Another issue is that the manuscript attributes the model-observation GPP mismatch to excessive nitrogen limitation without exploration of other potential drivers. The nitrogen manipulation experiments convincingly demonstrate that shrub productivity is highly sensitive to nitrogen availability within the model framework. However, they do not necessarily demonstrate that the

nitrogen cycle representation itself is the primary source of model bias. Arctic shrub productivity is strongly influenced by other factors such as snow conditions, soil temperature and moisture, and active layer dynamics, all of which are represented in QUINCY and can influence nitrogen availability through their effects on mineralization and plant uptake. Thus, an improvement in productivity after removing nitrogen limitation does not necessarily establish that nitrogen cycling is being represented incorrectly, but rather that nitrogen availability is constraining productivity in the model. Further evidence is needed to determine whether the observed model biases arise primarily from nitrogen cycle representation itself or from other simulated processes that regulate nitrogen availability. Given the importance of snow and soil thermal/hydrological conditions for Arctic shrub productivity, it would be useful to include at least a limited evaluation of simulated snow cover, soil temperature, and/or soil moisture at sites where observations are available. This would substantially strengthen the conclusion that nitrogen limitation is the primary factor controlling the model biases, in addition to providing further validation that the model is representing the sites accurately (beyond GPP).

We agree that this aspect on the model limitation by the nitrogen dynamics is overextended. The C-only model simulates up to 50 % higher GPP than the CN model configuration which better agrees with the 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.') accordingly:
The model-measurement agreement is lower at sites with local heterogeneous vegetation which is not captured by the model hydrology or represented by the shrub plant functional type.
To further support our hypothesis we extracted 89 measurements of leaf C/N ratio from Betula nana and 36 measurements from Betula glandulosa from the TRY database (Kattge et al. 2020). The C/N ratio estimated by the C-only configuration estimated at 7 stations (with deciduous species) 24.0 (median) agrees much better with the median 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 of a too strong N limitation in the model (CN configuration). We will change line 310 ff. in the manuscript accordingly (The entire CN vs. C-only discussion will be shifted to section 3.3).

3. Related to this, the model evaluation efforts rely primarily on seasonal GPP comparisons against field measurements and MODIS products. While these comparisons are useful, many of the newly introduced shrub parameter values affect structural and allocation-related traits (e.g., rooting depth, allometry, sapwood turnover, biomass partitioning). The current evaluation provides confidence in simulated GPP but less confidence in other modeled ecosystem characteristics that are important for the new shrub implementation. I would encourage the authors to discuss the limitations of a GPP-only evaluation and, if possible, compare modeled biomass, carbon stocks, phenology, vegetation structure, or nitrogen-related variables against observations.

We thank the reviewer for this constructive comment and acknowledge your suggestions for further validation. In addition to the comparison of measured monthly GPP from the most recent Arctic Carbon dataset ABCflux v2 with the model data as a critical benchmark for the shrub representation, we discuss the plausibility of the implemented shrubs in the manuscript text, e.g the biomass partitioning and its consistency with observed data (line 223-230). For the response to your comment 2 we retrieved additional field measurements of Arctic shrubs from the TRY database (Kattge et al. 2020). These data provides important insights but besides a notable time effort required for identification and data-handling also this comparison is limited by high variability across different local environmental conditions. We will explicitly discuss these limitations in the revised manuscript.

4. The manuscript would also benefit from a clearer description of the study sites. The authors mention that 16 Arctic sites are used throughout the analysis and repeatedly refer to specific sites, but there is no information about their geographic distribution, vegetation composition, climate characteristics, hydrological setting, etc. in the main text. Some general description of sites and the conditions/differences is needed in the methods section. The map and table that contain some site information are in Appendix A rather than in the main manuscript, which took me awhile to find (it's a bit confusing to have both an appendix and a supplement). Because site-specific differences are discussed throughout the Results and Discussion, it would be helpful to move the site map to the main text. Abbreviations for sites are also used throughout the manuscript and figures, but I could not find definitions for those abbreviations anywhere in the main text, which was confusing as a reader not familiar with these

sites. This could be resolved either by defining them somewhere in the main text or by moving Table A1 to the main text.

We add a description of the study sites in line 144:
The 16 study sites are located across the North US, Canada, Finland, Norway, Sweden and Siberia as seen in Figure 1 (map). These are wet, moist or dry tundra, fen, bog or boreal forest ecosystems vegetated by either a mixture of shrubs, trees or grasses or only shrubs (Tab. A2) .
We will extend this table with more information about the hydrological state of the sites. Additionally, we will move Figure A1 to section 2.4 and list the site abbreviations in the captions of Figure 1, 2 and 3.

5. Related to 4) above, the paper notes that several sites contain wet tundra, fen ecosystems, forests, or other heterogeneous vegetation that are not represented in the model. Because of this and because of the lack of description of the sites and their environmental conditions or gradients, it is difficult to separate potential model structural deficiencies from site representation errors. The authors should include some explanation of how many evaluation sites are considered representative of the simulated shrub PFTs.

Based on the ecosystem classification given in the metadata of the ABCflux data (extended table) 3 out of 16 sites (Siikaneva (SIL), Lettuoso (LET) and Stordalen (STR)) are not fully represented by the new shrub PFT. Siikaneva (SIL) and Lettuoso (LET) are classified with some fraction of boreal forest which is only represented by the default (tree) PFT. For parts of the measurements at the STR and SIL site the ecosystems are classified as peatland and fen ecosystem which the model hydrology does not capture yet. Accordingly, we will add to line 183:
Overall, 3 out of 16 sites (Siikaneva (SIL), Lettuoso (LET) and Stordalen (STR)) are considered as not representative of the simulated shrub PFTs due to their heterogeneous ecosystem.

6. It is somewhat unclear where some of the new parameter values for the two shrub PFTs come from. They are listed in table 2, but not all of them have references. Even for the ones that do have a reference listed, it would be helpful to know whether those measurements come from any the study sites, were measured somewhere else, taken from a database, etc. (given that the reader is likely not familiar with those studies). For that ones that do not have a reference, where did they come from? Where they tuned, or determined based on site knowledge, etc?

The one reference entry missing in table 2 for the parameter "Root distribution" will be filled with
determined based on tuning. The other entries will be extended as e.g. *measurements from TRY database (Kattge et al. 2020).*

7. The general readability / interpretability of the figures needs improvement (see technical corrections for more detail). The figures are often difficult to interpret due to the large number of site panels, use of site acronyms without definition, and limited contextual information. The authors should consider redesigning some figures to better highlight the primary findings and improve readability, and assess whether all the figures shown in the appendix are necessary to support the conclusions of the paper.

Please find the proposed changes for the figures below.

Technical corrections
As mentioned above, the figures need to be modified to be publication-ready, as many of them are currently difficult to read and interpret. Here are my specific suggestions that apply to most / all figures:
Units should be included in the figure caption as well as in the axis labels. Subscripts and superscripts need to be added where relevant (e.g., Figure 4, CO2), and the y axis labels all need superscripts in the labels. Please capitalize the axis title and make the axis labels larger than the numbers along the axes. The axis label should be centered if it applies to all subplots (e.g., in Figure 1). A “Site” x axis label is needed for several of the figures that list site acronyms on the x axis but have no axis label. The labels in the legends need to be defined in the caption in some cases (e.g., Figure 1, the legend acronyms are confusing). The study site map needs larger text and labels. For the supplement figures, the text is too small to read across the board (numbers, dates, etc. as well as axis labels). Figure S1 should have a y axis label across all subplots, rather than just on the top one.

If the axis labels are the same across all plots, the small labels on each sub panel should be replaced with one large label going across each axis (for example, left side and bottom of each plot for y and x axes, respectively). The legend and site labels are also too small to read. Figure S9 needs an x axis label, as do S10, S12, S17.

Thank you for the detailed suggestions. We will add the definition of the site acronyms to the caption of Figure 1-3, each. The caption of Figure 1 also has to be harmonized with the figure labels. Figure 2 and 3 will get a x-axis label.
We will introduce superscripts to the legend and the y-axis of Figure 4.
We will add the units of figure 6 in the caption.
We will redesign the site map and shift it to the main text. Figure A2 will be further moved to the supplement. We will re-do the supplement figures improving axes labels and font size as well as removing subpanels which are not used e.g. Figure S7a. Also Table S1 can be removed.

There are various typos and readability issues that should be addressed throughout the manuscript. I noted some of them below: -line 30 – "short statue" should be changed to "short stature"
-lines 31-33: This sentence doesn't make sense grammatically and needs to be revised. For example, "Importantly, shrubs have faster wood production than trees and therefore have a 20-50% shorter growing season (refs). This is a key adaptation to Arctic conditions, including. . . "
-lines 33: "evergreen and deciduous" should be replaced with "evergreen or deciduous". I would also remove "species" here as evergreen and deciduous are not species, but rather broader ways to group plants/species.
-line 47: replace "rising" with "increasing"
-line 57: "encouragement" should be "encroachment". I'm not sure what "shrub encroachment can not only largely change with climate" means, please clarify
-line 272: change "foilage" to "foliage"

We will carefully tackle these typos in the revision. Line 31-33 is reformulated to:
- Line 30: we will change as proposed
-Line 31-33 will be changed to *Shrubs exhibit faster wood production compared to trees, allowing them to complete their growth within a growing season that is 20–50% shorter (refs).*.
- Line 33: We will make proposed changes.
- Line 47: we will change accordingly.
-Line 57 is changed to: *Besides climate there are several other factors influencing shrub establishment and growing.* - line 272: we will correct the typo