

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 adapted the text in Sec. 3.1 and the abstract accordingly. We changed 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 shifted the discussion on the CN vs. C-only model configuration to Section 3.3 as suggested.

We adjusted 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).

65

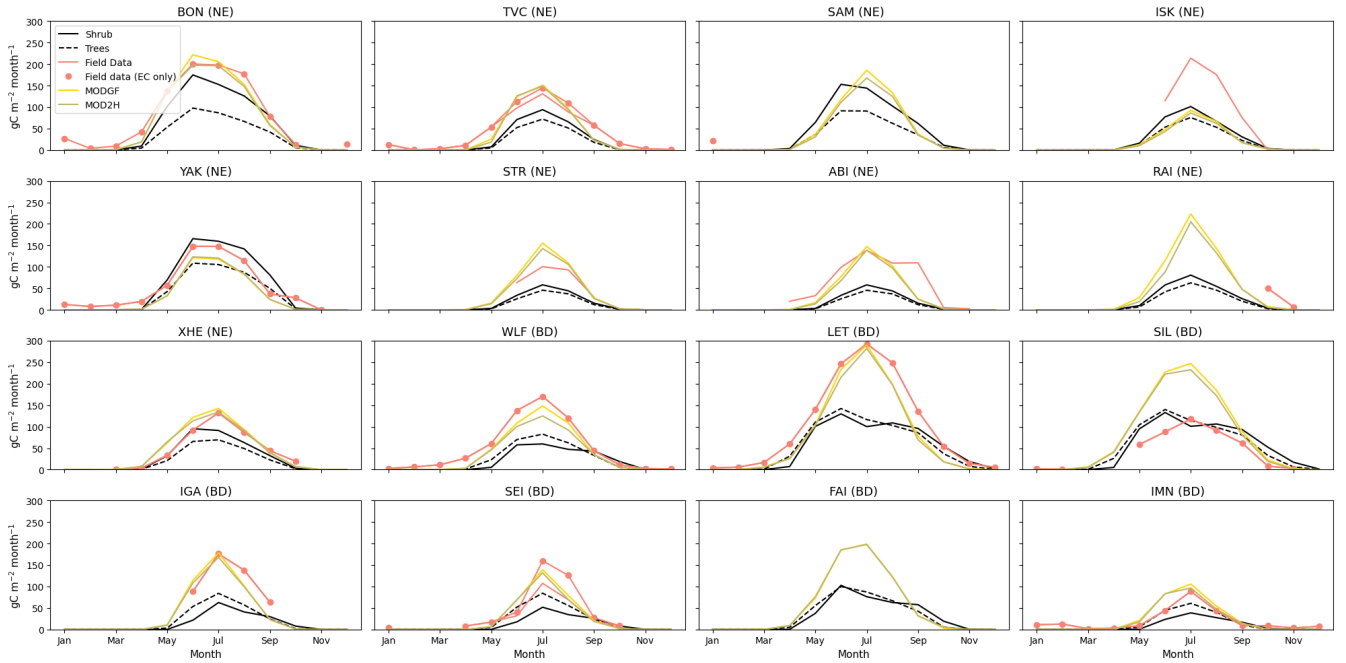


Figure 1. Multi-year mean seasonal cycle of GPP [gC m⁻² month⁻¹] by measurements (field data and MODIS satellite data: MOD17A2HGF and MODIS17A2H) and model with new shrub PFT, tree PFT and graminoid PFT at 16 sites (BON: Bonanza (US), TVC: Trail Valley Creek (CA), SAM: Sammallunturi fell (FI), ISK: Iskoras (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

70 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.

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

85 *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 improved 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.

110 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. Further, we evaluated the need of each supplement figure and re-structured the figures where necessary.

Table S1 can be removed as well because it is not referenced.

115 We apologize for two missing legend entries in S14, these were added.

TECHNICAL COMMENTS:

120 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 125 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.

135

We checked and improved the citations, grammar and punctuation.

We apologize for missing consistency among the figure colors and the site lists. We harmonized that.

Also, we agree to shift Figure A1 to the main text.

We harmonized the order of the site lists and used color bars.

140 We corrected the sentence meaning in Lines 205-206. Indeed, it compares evergreen shrubs with the default (tree) PFT.

We improved 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.

Reply to Reviewer comment 2

General comments

145

...

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 parameterizations; 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 150 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.

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

Specific Comments

165 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 followed 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:
175 *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:
205 *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 sup-

porting our hypothesis of a too strong N limitation in the model (CN configuration). We changed line 310 ff. in the manuscript accordingly (The entire CN vs. C-only discussion was 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 explicitly discussed 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 extended this table with more information about the hydrological state of the sites. Additionally, we moved 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 (Siikanen (SIL), Lettuoso (LET) and Stordalen (STR)) are not fully represented by the new shrub PFT. Siikanen (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 added to line 183:

Overall, 3 out of 16 sites (Siikanen (SIL), Lettuoso (LET) and Stordalen (STR)) are considered as not representative of the

simulated shrub PFTs due to their heterogeneous ecosystem.

265 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?

270 The reference entry missing in table 2 for the parameter "Root distribution" was filled with *determined based on calculations*. The other entries which miss a citation stem from the TRY database and was extended as e.g. *measurements from TRY database (Kattge et al. 2020)*.

275 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.

280 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:

285 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. 290 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.

295 Thank you for the detailed suggestions. We added 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 got a x-axis label.

We introduced superscripts to the legend and the y-axis of Figure 4.

We added the units of figure 6 in the caption.

300 We redesigned the site map and shifted it to the main text. Figure A2 was moved to the supplement. We re-created the supplement figures improving axes labels and font size as well as removing subpanels which are not used. 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"

305 -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"

- 310 -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 carefully tackled these typos in the revision. Line 31-33 is reformulated to:

- 315 - Line 30: we changed as proposed
-Line 31-33 was 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 did proposed changes.
- Line 47: we changed this accordingly.
- 320 -Line 57 is changed to: *Besides climate there are several other factors influencing shrub establishment and growing.* - line 272: we corrected the typo