

Responses to Reviewer#1's comments:

We thank Reviewer #1 for the insightful and constructive comments. We also appreciate the time taken to review this revision after a considerable delay, which was due to the author being on parental leave.

We have carefully addressed the comments to improve the clarity, precision, and overall quality of the manuscript. Detailed, point-by-point responses are provided below. Nearly all suggestions have been incorporated, and the corresponding revisions are highlighted in the manuscript. For a small number of comments where changes were not made, we provide reasoned explanations in this response document.

Line numbers refer to the revised version.

Specific comments:

Line 107-117: Both hypotheses and research questions are stated. Both are clear, concise and relate well to what the reader can expect from the paper. However, maybe one (either hypotheses or questions) would suffice? This is of course just a matter of taste, and not of scientific importance, but the text would likely be made even more concise by choosing only one of them.

We thank the reviewer for this insightful comment. While we agree that presenting either hypotheses or research questions alone could improve conciseness, we chose to retain both to clearly distinguish between the conceptual framework (hypotheses) and the specific analytical focus (research questions). Minor revisions have been made to reduce overlap and improve clarity.

Equation 1: Here, perhaps you could just show the entire equation for VPD rather than VPsat (since VPD is the variable used throughout the analyses)? You also refer to this equation on line 205 as an equation for calculating VPD. I suggest writing it as " $VPD = (0.6108 \times e^{((17.27 \times T_a)/(T_a + 237.3))} - VPa$ ", and explain that the formula in parentheses refer to VPsat. Alternatively adjusting the text on line 205 to clarify that Eq. (1) does not calculate VPD.

Thank you for this helpful suggestion. We have revised the equation to explicitly present VPD and clarified the description in the text to ensure consistency.

Line 153-170: This section is somewhat confusing regarding detrending of data. It seems here like an intrinsic part of GI is the fact that the data has been standardized to remove time-dependent factors. However, in your statistical analyses you explain that GI data is used both as detrended and non-detrended. Is the GI simply radial increments (as stated in the 2nd paragraph), in which case it is directly comparable to the data collected from the six experimental sites? Or is it already manipulated, in which case detrending the GI might be a case of "double-detrending" of data? This could be clarified, perhaps through a thorough explanation of how GI is calculated.

Thank you for pointing out this lack of clarity. The Growth Index (GI) is not a simple measure of raw radial increment; rather, it is standardized to remove the effects of age, site conditions, and competition. The normalization of GI sets the mean to 100 over the overall study period,

with deviations around this value representing variability likely associated with environmental factors across multiple temporal scales.

To address the concern regarding potential ‘double detrending’, we clarify that additional detrending was applied to both GI and climate variables only to remove low-frequency variability. This step necessary to avoid inflating explained variance (R^2) due to shared long-term trends, as our main goal in this analysis was to assess interannual variability, not the full relationship across all timescales. However, we do later explore the full relationships. By removing low-frequency components, we focus on high-frequency signals that better reflect year-to-year climate-growth dynamics.

Line 175: The correlation is between plot-level BAI and NFI GI. It is not entirely clear from this section if the correlations were carried out on detrended or non-detrended data. This could be added for clarity.

Thank you for noting this. We have clarified that the correlations presented in Figure S2 were computed using non-detrended data.

Line 224: There are some (although very few) series that have low correlations to the master chronology. Were these removed prior to analyses or kept and trusted? The low number of low correlations mean that they likely do not influence the results much, but a note on how they have been handled could be added for transparency.

All series, including those with lower correlations to the master chronology, were retained in the analyses. Given the relatively short time series and the influence of natural forest dynamics, we found no evidence suggesting cross-dating errors. As pointed by the Reviewer, we consider that the low number of such cases also suggests a negligible influence on the overall results. We have added this clarification to the revised manuscript to improve transparency (see Line 240).

Comment: Regarding the Carbon isotope analysis: I am unfamiliar with methods regarding this, and can thereby not criticize the method. However, the analysis seems to have been carried out by a reliable laboratory and I see no reason to think that the data is of poor quality.

We have clarified the methodological description in the revised manuscript (see section 2.3.6) to specify that the $\delta^{13}\text{C}$ of α -cellulose samples were measured using an isotope ratio mass spectrometer (IRMS) at the Stable Isotope Laboratory (SILL), Natural Resources Institute Finland (Luke), a well-established facility with extensive experience in stable isotope analyses.

Line 313: According to the results you seem to have done 3-month averages rather than “each site and month” which seems to be only presented in supplementary results. This could be clarified.

We have clarified that climate variables were analyzed as 3-month averages in the main analysis (Figure 5), while monthly results are presented in the Supplementary Materials (Figure S17). Additionally we added a 3 month period analysis in (Figure S18) The text now specifies when each temporal aggregation is applied and also if it contains detrending or not, while retaining a general description where appropriate for readability.

Line 319: Here you use MJJ while you present climatic trends of JJA (or August) in Figure 1. Is there a reason why MJJ was chosen? Furthermore, coherency between Figure 1 and this choice could provide the reader with a better understanding of the input to the analysis.

We appreciate this observation. The period MJJ was selected because it best captures the climatic conditions most relevant to the physiological processes represented in LMM analysis. In contrast, Figure 1 presents JJA trends to provide a broader seasonal context. While aligning these periods could improve visual coherence, the distinction reflects their different purposes: Figure 1 provides general climatic context, whereas the LMM focuses on the biologically relevant time window. We have clarified the use of the MJJ period in the LMM description in the revised manuscript to avoid confusion and improve coherence (see Lines 345-350).

Figure 2: There is a drastic decline in data towards the later years (as shown in the lower panel) –Why is that and doesn't it risk skewing your interpretation of the temporal development of growth? If you agree that the decreasing data quantity could influence the results, I suggest adding a caveat in the text about this.

The decline in GI sample size in the most recent years is primarily due to limitations in available NFI material and the cumulative nature of the tree-core sampling design, whereby earlier periods contain more observations.

We acknowledge that the reduced sample depth toward the end of the GI time series may influence the representation of recent growth variability. To address this, we have now estimated 95% confidence intervals for the GIs, which are presented in Appendix A. This provides an explicit assessment of uncertainty, particularly for the most recent years, and is reflected in the widening confidence intervals in that period.

Importantly, the overall temporal patterns remain consistent with our BAI plot-based measurements, which provide independent support for the observed trends. We have also added a note in the revised manuscript to clarify the decreasing sample size toward the end of the period and to advise cautious interpretation of the most recent values. (Line 454, and Appendix A)

Table 1 and 2: The post-hoc tests that the significance levels are based on don't seem to be mentioned in the methods. These could be added so that they don't appear in the results for the first time. Further, I personally do not agree with using p-value adjustments as this can potentially force a Type II error onto data that otherwise reveal clear differences. However, this is a personal opinion, and p-value corrections are very common, so I do not expect you to change this – I just wanted to air my opinion on the matter.

We have now specified in the new section 2.4.1. within the Statistical methods that significance levels are based on Bonferroni-adjusted pairwise comparisons. While we acknowledge that such adjustments may increase the risk of Type II errors, we retained this conservative approach in line with common statistical practice.

Line 375: You write that the decline of Norway spruce BAI was especially large under dry conditions. Considering Figure 3B, I'm not sure I agree. The decline seems to be the same from 2015 in both soil moisture types. If anything, the relative decline seems greater under

wet conditions (e.g. by looking at the mean average lines: $0.4/0.8 = 0.5$ versus $0.6/1.0 = 0.6$, when comparing the last data entry to that of 2015). If you still think that the decline is larger under dry conditions, the reasoning behind this could be clarified.

We agree that the visual decline in Norway spruce BAI during recent years appears similar under both dry and wet conditions in Figure 3B. Our original interpretation was primarily based on visual inspection of the figure.

Revised breakpoint analyses identified a consistent change on the mean and slope around 2016 for Norway spruce variables (i.e., GI, BAI, and iWUE), whereas Scots pine showed an earlier breakpoint around 2009 only for BAI. To facilitate consistent comparisons among soils and species, a harmonised breakpoint around 2016 was subsequently applied in the LMM analyses.

To provide stronger statistical support for the interpretations presented in Figures 2 and 3, we have now added supplementary Tables S13–S18, which report temporal slopes and estimated marginal means (EMMs) for GI, BAI, and iWUE across species, soils, and pre-/post-breakpoint periods. All corresponding result descriptions in the manuscript were subsequently reviewed and revised based on these statistical outputs.

Specifically for Norway spruce BAI (Tables S15–S16), post-2016 temporal trends were negative under both dry and wet conditions (dry: $-0.0691 \pm 0.0145 \text{ yr}^{-1}$; wet: $-0.0605 \pm 0.0145 \text{ yr}^{-1}$), while the difference in slopes between moisture classes was not statistically significant ($\Delta\text{slope} = -0.0086 \pm 0.0205 \text{ yr}^{-1}$, $p = 0.683$), we no longer state that the decline was clearly larger under dry conditions. Instead, the revised text emphasises that both moisture classes experienced substantial post-2016 declines, while dry plots maintained slightly higher BAI values overall during the post-change period according to EMM estimates (Table S16).

Line 388: Regarding the GI and drier conditions. You mention here that trend changes in GI were more pronounced in drier conditions. However, can the comparison between drained organic soils and mineral soils be considered as a comparison between “wet” and “dry” soils? You introduce the soil types in the paragraph around L88-L100. However, it is not clear from this that organic soils (drained or not) should be considered wetter than mineral soils. I suggest clarifying that the trend changes were more pronounced on mineral soils – not at sites with drier conditions.

We agree that drained organic soils should not be interpreted as a direct proxy for ‘wet’ conditions, nor mineral soils as inherently ‘dry’.

We have restructured the results section referred to section 3.3, our intention was not to contrast mineral and organic soils as dry versus wet environments, but rather to describe overall temporal patterns in GIs for both soil types in Scots pine and Norway spruce. This provides a broader context for their growth responses across southern Finland and serves as a link to the study-site-level information, where we assess differences in ‘dry–wet’ soil moisture conditions for each species and do not find differences that are statistically robust.

Specifically, we have revised the Results associated with Figure 4 to ensure that the description focuses on species-specific temporal trajectories and general growth trends. These revisions should eliminate confusion between soil type and moisture condition and better align the text with the scope of our analyses.

Line 388: Regarding the BAI and drier conditions. According to Table S13, only 2 of 6 time series changed significantly, and 1 of those was from a wet site (not counting the wet site with a significant change in slope but without a positive-to-negative trend change). While the change in the drier site indeed was more negative, the total slope change was greater in the wet site. Therefore, while I think you can argue for the fact that the drier sites show more significant trend changes, you could also argue for the opposite. Hence, I would like to see a clarification on what you base that “BAI [...] were more pronounced at sites with drier conditions” on. Figure 4 (figure text): “In contrast, the coloured lines and dashed lines highlight time series in which the segmented regression method detected one or more change points.” – It is not clear from the text what the difference is between the coloured solid and coloured dashed lines are. I suggest clarifying (if I have understood it correctly) that the coloured solid lines represent time series with one or more change points detected and that coloured dashed lines represent the model output from the segmented regression analysis.

We agree that the relationship between soil moisture conditions and the magnitude of BAI changes is not fully consistent across all sites and species. As it is correctly pointed out, both dry and wet sites can exhibit strong post-breakpoint declines, and in some cases wet sites showed larger absolute slope changes.

Our original statement was primarily based on visual inspection of the site-level trajectories and the observation that some Scots pine sites under drier conditions exhibited earlier or clearer breakpoint detections. However, we acknowledge that this wording overgeneralised the role of soil moisture conditions.

To provide stronger statistical support for these interpretations, we have now added supplementary Tables S13–S18, reporting temporal slopes and EMMs derived from the breakpoint-guided analyses for GI, BAI, and iWUE. Based on these analyses, we revised the manuscript text to avoid implying a consistently stronger decline under dry conditions.

Regarding Figure 4, we agree that the distinction between coloured solid and dashed lines was insufficiently explained. We have therefore revised the caption to clarify that grey solid lines represent observed time series without statistically significant breakpoints, coloured solid lines represent observed time series with one or more detected breakpoints, and coloured dashed lines represent the fitted segmented regression models for those breakpoint-detected series.

Line 396-407: It is not clear whether these results are based on detrended or non-detrended data. In the methods you mention that analyses are carried out on both. This could be clarified, even if the results are similar independent of detrending.

We have clarified in the caption that the results presented in Figure 5 are based on non-detrended data.

Line 466: “Norway spruce [...] was more severely impacted by rising VPD” – this can be inferred from the analysis of pre- and post-2015. However, the overall regression analyses seem to complicate the findings somewhat. Except for previous year autumn VPD, there doesn’t seem to be much that suggest that Norway spruce is more sensitive to VPD (in fact, current year summer months’ VPD suggest a weakly opposite pattern). This complexity could be elaborated in this section.

Apparent differences between the correlation analyses (Figure 5) and the linear mixed-effects models (Figure 6) arise from differences in temporal structure and analytical design.

The correlations in Figure 5 are based on the full time series (1990–2023) and therefore represent average relationships across the entire study period, without accounting for potential temporal shifts between pre- and post-change periods (1990-2015 and 2016-2023, respectively). In contrast, the LMM approach explicitly separates these periods, allowing changes in climate–growth relationships over time to be detected.

This difference in temporal aggregation likely contributes to the weaker and more variable correlations observed for BAI, which integrate multiple co-limiting factors (e.g. site fertility and stand conditions) when assessed over the full period. We have updated in the revised manuscript the more consistent lagged and seasonal sensitivity of BAI to VPD when temporal structure is accounted for (see Lines 550-580).

Line 473: “Recent studies further indicate that the degree of atmospheric drying registered over the last 400 years is unprecedented” – This makes it seem as though the entire 400 years are unprecedented. I suggest rephrasing to something like: “Recent studies further indicate that the current degree of atmospheric drying is unprecedented compared to the last 400 years”

The sentence has been rephrased in the revised manuscript.

Line 538: This part of the discussion is very nicely written and highly relevant to forest management, and most of the suggestions seem intuitive. However, I’m not sure I understand why structural diversity would lead to alleviated water stress. Could this be explained or expanded on?

We have clarified the role of structural diversity in alleviating water stress by explicitly mentioning variation in rooting depth, in addition to tree ages and sizes. This clarification is supported by a new reference (Hanby et al. 2025) included in the revised manuscript.

Galen Hanby, Omkar Joshi, Lu Zhai (2025) Drought impact on tree productivity: Varying roles of tree size and structural diversity in 18 woody species along gradients of slow-fast growth strategies, Agricultural and Forest Meteorology, 369,110592, <https://doi.org/10.1016/j.agrformet.2025.110592>.

Line 551: Same here. It is not intuitive to me why CCF would reduce intra-stand competition for water.

We have clarified the mechanism by which silvicultural adjustments can reduce intra-stand competition for water (see Lines 685). This phrasing emphasizes how structural heterogeneity in the stand helps distribute water demand and alleviates competition among trees.

Line 549: Why is this only imperative in drained organic soils? The paragraph starts by showing that well- and poorly drained soils both have marked declines in growth. And considering Figure 2B, it seems like the mineral soil, rather than the drained organic soil, is more sensitive.

In this section, we specifically focus on Norway spruce growing on drained organic soils under different soil moisture conditions.

Suggestions on spelling and sentence structure:

All proposed edits regarding wording, grammar, and consistency have been carefully implemented throughout the revised manuscript.

In **BOLD** are suggestions for added words,

~~STRIKETHROUGH~~ are suggestions for removed words.

Line 19: “increased more steeply **in** Scots pine”

Done.

Line 41: “These droughts, often coinciding with elevated air temperatures (Ta), increased vapor pressure deficit (VPD) and reduced soil moisture, are likely to become more frequent under continued warming”

Done.

Line 74: “reflects a species’ ability” – apostrophe.

Done.

Line 90: “Drainage induced lowering **of** the water table”

Done.

Line 117: “How **does** Ta, VPD and SPEI”

We thank the reviewer for this suggestion. However, we have retained “How do...” because the subject of the sentence, Ta, VPD, and SPEI, is plural. Therefore, the original phrasing is grammatically correct and has not been modified.

Line 126: “raw material for **the** forestry sector”

Done.

Line 149: “0.05 °C” – degree sign

Done.

Line 166: “Tree-ring cores were collected at ~~diameter at the~~ breast height”

Done.

Line 191: “the plot. **In** Scots pine” – capital letter

Done.

Line 204: “we obtained the mean daily Ta and VPDair from the observational data interpolated to a 10 km × 10 km grid covering Finland ~~obtained~~ from the Finnish Meteorological Institute”

Done.

Line 219: “Each tree core was immersed in water for at least 30 minutes to avoid breaking the cores, when mounted on wooden supports, and ~~prepared~~ the core surface **was prepared** with a microtome”

Done.

Line 301: “(Killick, 2011). This” – dot

Done.

Line 310-316: Present and past tense: “we use” and “we fit” compared with “Detrending was” and “We applied”.

Done.

Line 375: “making the growth plateau~~d~~ and decline”

Done.

Line 383: “is shown in more **detail** in Figures S15–S16.”

Done.

Line 398: “[...] y climate variables, ~~including~~ Ta, SPEI, and VPD” – “including” suggests that there are other variables as well.

Done.

Line 412: “Growth sensitivities ~~to SPEI~~ remain close to zero for SPEI and are negative for VPD”

Done.

Line 466: “Norway spruce showed a notable decline in growth, particularly after 2015, and was more severely impacted by rising VPD, ~~this~~. **This** response is likely attributed to its physiological characteristics and adaptation to colder, more humid conditions, resulting in a sharp decline in growth due to increasing water stress”

Done.

Line 540: “. In contrast, [...]” – dot

Done.

Line 566: “Scots pine exhibited greater drought vulnerability on organic soils, **likely** due to limited aeration and impaired rooting conditions during dry periods” – as it is now, it sounds as though the mechanistic pathway of impaired rooting conditions was tested. I suggest adding some cautious wording.

Done.

Responses to Reviewer #2's comments:

I thank the authors for their thoughtfully written paper and for EGU's BG journal and community for the opportunity to review this manuscript. This manuscript investigates long-term growth and eco-physiological responses of Norway spruce and Scots pine in southern Finland, using Growth Indices (GI) alongside basal area increment (BAI), intrinsic water-use efficiency (iWUE), and carbon isotope discrimination ($\delta^{13}\text{C}$) using Finnish NFI as well as independent observational data. The authors aim to identify temporal trends in growth and to attribute species-specific differences in these trajectories to underlying physiological responses to increasing water limitation. The study addresses an important question in forest ecology and climate change, and the integration of NFI growth metrics with independent field observations and isotopic indicators is a clear strength. This study is well suited for the journal and will be of particular interest to the readership of EGU Biogeosciences. However, several aspects of the statistical treatment and interpretation (particularly related to uncertainty, sampling imbalance, and effect-size interpretation) require clarification or revision before the conclusions can be considered robust.

We sincerely thank the Reviewer for their thoughtful and thorough evaluation of our manuscript. We appreciate the positive remarks regarding the integration of NFI growth metrics with independent field observations and isotopic indicators, as well as the relevance of our study to forest ecology and climate change. We also acknowledge the Reviewer's constructive feedback regarding aspects of statistical treatment, interpretation of uncertainty, sampling imbalance, and effect-size considerations. In the revised manuscript, we have thoroughly addressed these points to enhance the clarity, transparency, and robustness of our analyses and conclusions. Detailed, point-by-point responses are provided below. We have incorporated nearly all the reviewers' suggestions, and the corresponding revisions are highlighted in the manuscript. For a small number of the comments where changes were not made, we have provided reasoned explanations in this response document.

We also appreciate the time taken to review this revision after a considerable delay, which was due to the author being on parental leave.

Line numbers refer to the revised version.

Specific comments:

Figure 1: This is an effective figure that clearly illustrates the increasing vulnerability of the study region to climate change. I have one stylistic suggestion: for panels in which the color scale includes zero (Fig. 1B, 1E), a diverging color scheme may be more appropriate. In the current perceptually uniform, non-diverging scheme, values near zero are difficult to distinguish from potentially meaningful positive or negative deviations.

We agree that a diverging colour scheme can improve the visibility of values near zero. However, in the revised manuscript, applying a diverging colour scale to panels 1B and 1E did not fit well with the intended message of these maps. These panels are designed to highlight that trends in the south are steeper than in the rest of the region, rather than to emphasize opposite trends. Therefore, we chose to retain the current perceptually uniform colour scheme, as it more clearly represents the spatial gradient in trend magnitude while still

allowing small positive and negative deviations to be distinguished, and preserves the overall interpretability of the figure

Equation 1: In the text (e.g., line 206), Equation 1 is referenced as the calculation for VPD, but the equation itself is labeled as VP_{sat} , which is a distinct quantity. This discrepancy should be clarified here and in line 206.

We thank the reviewer for this comment. As previously noted by Reviewer #1, we have clarified the text and the Eq. (1) to avoid confusion. The text has been updated to indicate that the expression in parentheses corresponds to saturation vapor pressure (VP_{sat}), while the full equation calculates VPD. This ensures consistency between the equation label and its description in the revised manuscript.

Lines 153–170: Figure 2 shows a substantial decline in the number of sampled trees after approximately 2005. This section of the Methods (“Finnish NFI data”) should acknowledge and explain this pattern. Additional context on the history and design of this dataset may be warranted, particularly given the apparent temporal and spatial unevenness in sampling effort (see comments below). Further comments later in the methods should explain how this is addressed to prevent statistical artifacts.

We thank the reviewer for highlighting this important aspect of the NFI dataset. We have added clarification in the section 2.2. to acknowledge the decline in the number of sampled trees after ~2005, which primarily reflects the structure and temporal design of the NFI, as well as limitations in available historical plot data (Lines 160)

We also acknowledge that the reduced sample depth toward the end of the GI time series may influence the representation of recent growth variability. To address this, we have now estimated 95% confidence intervals for the GIs, which are presented in Appendix A. This provides an explicit assessment of uncertainty, particularly for the most recent years, and is reflected in the widening confidence intervals in that period.

Line 173: These sites are described as “experimental,” but the study is observational and does not involve experimental manipulation. I suggest rephrasing this terminology accordingly.

We have revised the terminology throughout the manuscript to refer to these locations as ‘study sites’ rather than ‘experimental sites’ to accurately reflect the observational nature of the study. This change ensures clarity and consistency with the study design, which does not involve experimental manipulation.

Line 320: Were predictors mean-centered only ($y - \bar{y}$), or standardized using z-scores ($(y - \bar{y})/sd(y)$)? It would be nice to be extremely clear here since this is a decision that meaningfully impact the interpretation of results presented later.

Predictors were mean-centred, not standardized. We used mean-centring to improve the interpretation of model intercepts and interaction terms while retaining the original units of the climate variables. Thus, model coefficients remain directly interpretable in ecological terms, for example per unit change in VPD or SPEI.

To avoid ambiguity, we have revised the Methods section to explicitly state that climate predictors and years were centred, but not standardized, before fitting the linear mixed-effects models (Lines 350-355)

Lines 335–345: Figure 2 indicates a marked decline in the number of sampled trees after approximately 2005, as well as a strong imbalance in sampling effort between soil types (with roughly five times more trees sampled on mineral soils than on organic soils). However, the plots in Figure 2 (and the analyses in Section 3.1 that rely on them) appear to ignore this fact, or otherwise be based on simple annual mean estimates of GI.

This raises statistical concerns. NFI data are inherently samples from a population, and the authors here treat annual estimates as unbiased and known values. Collapsing the data to annual means implicitly treats these values as known without error, thereby ignoring uncertainty arising from finite and declining sample sizes. This issue becomes increasingly severe in later years (e.g., post-2010), when sample sizes are smallest and sampling bias is most likely. Consequently, inter-annual variability and heterogeneous uncertainty among sites are likely underestimated, and differences between soil types may be overstated.

Moreover, and perhaps most importantly, because these periods of reduced and imbalanced sampling coincide with the most recent portion of the time series (i.e. the part of the time series where change point detection algorithms suggested a break point), they may exert disproportionate leverage on trend estimation and change-point detection, potentially leading to spurious or overconfident inferences. Under these conditions, the change-point analyses and associated p-values are difficult to interpret as valid. Explicitly accounting for sampling uncertainty—e.g., through hierarchical modeling or another form of uncertainty propagation—would substantially strengthen the robustness of the results.

At a minimum, the authors should provide uncertainty estimates around these trends, visualize this uncertainty in the time-series plots, and propagate uncertainty from the fitted models into the subsequent change-point analyses rather than relying solely on point estimates.

We agree that the decline in sample size over time and the imbalance between soil types are important considerations when interpreting the NFI-based GIs, particularly in the most recent years.

The observed decrease in sample size after approximately 2005 reflects the structure and temporal design of the Finnish NFI including its rotating sampling scheme and the availability of tree-core data. As a result, earlier periods accumulate a larger number of observations than more recent years, and mineral soils are more extensively represented than organic soils.

We acknowledge that treating annual GI values as point estimates without explicitly visualising their uncertainty can lead to underrepresentation of variability, particularly in periods with lower sample depth. To address this, we have now estimated 95% confidence intervals for the GIs, which are presented in Appendix A and reflected in the revised manuscript. These confidence intervals explicitly account for sampling variability and show increasing uncertainty toward the most recent years, consistent with the declining sample size.

Importantly, the GI estimates are derived using mixed-effects models following established NFI methodology (Henttonen, 2000), which partially accounts for hierarchical structure and variability in the data. In addition, the temporal patterns observed in the NFI-based GIs are consistent with independent plot-based BAI measurements, providing external support for the robustness of the main trends (Figure S2).

Regarding the potential influence of uneven sampling on breakpoint detection, we agree that reduced sample size in recent years may increase uncertainty around the exact timing and magnitude of detected breakpoints. We have clarified this limitation in the revised manuscript (see Lines 390–395). However, the convergence of evidence between GI and BAI, but also with $\Delta^{13}\text{C}$ and iWUE, supports the conclusion that a substantive shift in growth and physiological dynamics occurred, rather than this pattern arising solely as an artefact of the sampling structure.

We have revised the manuscript to i) explicitly describe the temporal and spatial sampling structure of the NFI data (see Lines 165–178), ii) include uncertainty estimates for GI time series (see Appendix A), and iii) clarify the limitations associated with declining sample size and their implications for interpretation of recent trends and breakpoint analyses (see Lines 392–395 and 455–456).

Lines 413–422 and Figure 6: The Results section would benefit from a clearer discussion of the biological relevance of the estimated effect sizes. For example, given the fitted models, what does a one-"unit change" (as written in Figure 6) in VPD or SPEI correspond to in terms of changes in BAI or $\delta^{13}\text{C}$, and are these effects large, moderate, or negligible from a biological perspective?

We have now expanded the Results section to clarify the biological relevance of the estimated effect sizes. Specifically, we now express regression slopes in terms of the expected change in BAI ($\text{m}^2 \text{ha}^{-1} \text{yr}^{-1}$) and $\Delta^{13}\text{C}$ (‰) per unit change in VPD or SPEI, and relate these values to the observed variability of each response variable. This allows effect magnitudes to be interpreted as proportional changes relative to site-level variability.

To further improve clarity, we now explicitly classify effects as negligible (<10% of observed range), moderate (10–50%), or large (>50%), consistent across both response variables. These classifications are summarized in the newly added Table 4, which compiles all slope estimates and their corresponding biological effect-size categories for both species, periods, and VPD and SPEI climate drivers.

Together, these revisions allow readers to directly assess whether climate-driven changes in BAI and $\Delta^{13}\text{C}$ represent negligible, moderate, or large shifts relative to observed ecological variability, rather than interpreting model coefficients in isolation.

Lines 465–470: Given the concerns about potential statistical artifacts in Figure 2 (see above), together with the relatively modest and somewhat opaque regression results in Figures 5 and 6, the evidence presented here that Norway spruce has been more strongly impacted by rising VPD appears limited. While the alignment with independent plot-level data and previous studies (e.g., Lagergren and Lindroth, 2002; Lévesque et al., 2013) is noted, this section as currently written provides only weak additional support beyond existing literature.

Additionally, this section appears to rely heavily on statistical significance, with relatively little discussion of the biological significance of the estimated effects (see also my previous comment). With p-values and R^2 values reported but limited interpretation of effect magnitudes, it is difficult for the reader to assess whether statistically significant results correspond to biologically meaningful changes.

We acknowledge the reviewer's concern that the evidence for Norway spruce being more strongly impacted by rising VPD requires careful interpretation, particularly given the potential limitations associated with NFI sampling (Figure 2) and the relatively modest effect sizes presented in Figures 5–6 and Table 4.

To strengthen the statistical support and transparency of our interpretations, we have now expanded the analytical framework presented in the manuscript. Revised breakpoint analyses identified a consistent change point around 2016 across multiple variables in Norway spruce, whereas Scots pine showed an earlier breakpoint around 2009 only for BAI. These breakpoint estimates were subsequently incorporated into linear mixed-effects models (LMMs) to evaluate temporal dynamics before and after the identified change periods. To facilitate consistent comparisons among soils and species, harmonised breakpoints were applied across the analyses.

In addition, we have added supplementary Tables S13–S18, which report temporal slopes and estimated marginal means (EMMs) for GI, BAI, and iWUE across species, soils, and pre-/post-breakpoint periods. These tables provide explicit statistical support for the interpretations associated with Figures 2 and 3 and allowed us to systematically reassess the wording throughout the Results and Discussion sections.

To address the reviewer's concerns more specifically, we revised the text to:

- 1. Contextualize statistical significance with biological relevance by discussing effect magnitudes in the original units of BAI and $\Delta^{13}\text{C}$, and by evaluating whether the observed responses are small, moderate, or large relative to typical interannual variability (see Table 4).*
- 2. We now, clarify the role of complementary datasets by emphasizing that regional-scale GI patterns, plot-level BAI& $\Delta^{13}\text{C}$ /iWUE trajectories, breakpoint analyses, and temporal trends collectively support the interpretation that Norway spruce experienced stronger recent growth constraints associated with increasing atmospheric drought conditions, together with a limited iWUE response.*
- 3. Additionally, we have refined the interpretation to avoid overstatement and to more clearly acknowledge dataset limitations. The revised text now presents the apparent VPD sensitivity of Norway spruce as a consistent but moderate pattern supported by several complementary analyses, rather than as conclusive evidence from any single dataset or statistical relationship.*

These revisions provide a more balanced interpretation integrating breakpoint analyses, temporal trend estimates, EMM comparisons, and biological significance by combining growth and iWUE responses, thereby improving the transparency and robustness of the conclusions.

Suggestions for spelling and grammar:

Note: I've removed comments here that were already pointed out by another referee.

Line 320: "Climate predictors and years were ****centered**** before"

Done.