Author response

How beech ecophysiology shapes temperate forest gross primary productivity – Part 2: Identifying critical timeframes across phenological stages

Jonathan Bitton[1], Bernard Heinesch[1], Catherine Charles[1] and Bernard Longdoz[1]

[1]Biosystems Dynamics and Exchanges, TERRA Teaching and Research Center, Gembloux Agro-Bio Tech, University of Liege, Gembloux, Belgium

Correspondence to: Jonathan Bitton (Jonathan.bitton@uliege.be)

Dear Editor and Referees,

We sincerely thank you for your constructive comments and suggestions, which helped us improve the quality and clarity of the manuscript. In the following pages, we provide a detailed response to all referee comments and indicate the corresponding changes made in the revised manuscript. The content is identical to that already provided in the author comments posted during the interactive discussion but has been reorganized to facilitate the assessment of manuscript changes.

For ease of consultation, this document is organized into three categories according to the extent of the resulting revisions: major changes, minor changes, and responses without manuscript changes. Within each category, comments are presented following the chronological order of the corresponding sections in the manuscript. Comments addressing the same section of the manuscript have been regrouped while preserving separate responses for each referee. For conciseness, figures have been omitted from the present document, as all updated figures are included in the revised manuscript.

Throughout the document, referees are identified using **R#** (e.g. R1, R2) followed by:

i. **C#**: general comment number, according to the structure adopted by the referee;
ii. **L#**: line-by-line comment referring to the corresponding line number;
iii. **Figure #**: comment related to a specific figure;
iv. **Table #**: comment related to a specific table.

For line-by-line, figure and table comments, numbering refers to the preprint version of the manuscript.

All modifications made in the manuscript are indicated at the end of each response and introduced by the symbol "⇨". References to sections, figures, and tables in the revised manuscript are highlighted in blue. Line numbers in blue refer to the tracked-changes version of the revised manuscript.

Sincerely,

The authors

Major changes

Allometric relationships and biomass estimation

<u>R1; L142</u>: "***What is the relationship between "trunk" and "shoot"? What data support the inclusion of root biomass in "B"?***"

We agree that the terminology used in the original version could be ambiguous. We have therefore adopted the more standard terms "aboveground" and "belowground" biomass. Aboveground biomass includes all aerial compartments (stem, branches and twigs), while belowground biomass refers to root biomass.

Total tree biomass (B) was derived from a site-specific allometric relationship, which directly estimates biomass as a function of tree circumference and explicitly includes both aboveground and belowground components.

Because the different biomass compartments follow similar power-law relationships with circumference, we retained a single total-biomass formulation rather than summing compartment-specific equations.

These relationships have been previously established for the Hesse site and validated. Their applicability is further supported here by comparison with independent allometric equations available for beech in the GlobAllomeTree database (Henry et al., 2013).

<u>R2; L142</u>: "***The reference to "Ottorini, personal communication" for the allometric equation is bit unusual. Is this equation available in any citable form?***"

We agree that citing a personal communication is not ideal. The allometric relationships used in this study are based on site-specific equations originally developed by Le Goff and Ottorini (2001) and later revised by J. M. Ottorini in 2012. While the updated formulation is not available as a standalone publication, it builds upon previously published relationships for the same site. To improve transparency, we provide additional details and justification for their applicability, including comparison with independent allometric relationships.

⇨ Changes in manuscript: We have:

i. Sect. 2.3.1 (L148–149): replaced the previous terminology describing tree compartments with the terms "aboveground" and "belowground" biomass;
ii. Sect. 2.3.1 (L149–150): revised the manuscript to explicitly cite Le Goff and Ottorini (2001) and clarified that the equations correspond to an updated version derived from these established formulations;
iii. Sect. 2.3.1 (L152–153); Supplement S2; Fig. S2: added a dedicated Supplementary section and figure describing the allometric relationships in detail and supporting their applicability at the study site.

Indicators definition and dependency on Part 1

<u>R1; C1</u>: "***Regarding your three GPP metrics: why these particular metrics? (Why not, for example, the rate of decline rather than the offset between two peaks)? Do you really need to name them "IRise," "IPeak," and "IDrop," rather than using short phrases to call them what they are? At the least, please include a detailed schematic labeling these metrics when they are introduced; I didn't understand what IDrop was until I made it to Figure 4, and even then, I still had some questions (is there always a second peak? What is the usefulness of this particular metric?).***"

We thank the referee for this comment, which helped us improve the clarity of the indicators.

Our choice of IRise, IPeak and IDrop was guided by the objective of capturing the three main, well-documented phases of seasonal GPP dynamics in temperate deciduous forests:

1. Early-season development (IRise): this phase is strongly linked to leaf ontogeny and reserve mobilization, both of which have been extensively studied in literature. The rate of increase in GPP provides a direct proxy for canopy development dynamics.
2. Peak productivity (IPeak): peak GPP reflects the period of maximum photosynthetic capacity, when environmental constraints are minimal and canopy structure is fully developed.
3. Mid-season decline (IDrop): this phase captures the reduction in GPP during summer, which is typically associated with increasing water and atmospheric stress. While we agree that this indicator is less intuitive, it was specifically designed to quantify the intensity of the decline relative to seasonal dynamics, rather than simply the senescence dynamics in autumn. This makes it particularly suited to identifying drought-related responses.

We acknowledge that IDrop was not sufficiently clear in the original version. The indicator does not rely on the systematic presence of a second peak in the GPP curve. Instead, it is derived from the wavelet representation of the signal, where the right-hand side of the wavelet centered on the GPP maximum captures the decline phase. The presence of a plateau or secondary peak only modulates the magnitude of the signal, without affecting the definition of the decline itself.

Regarding the number of indicators, we deliberately limited the analysis to three metrics to maintain clarity and avoid overcomplicating interpretations. Additional indicators, such as decline rate, were considered but not retained due to (1) weaker signal detectability, (2) higher uncertainty and (3) limited support in the literature.

Concerning terminology, we retained the IRise/IPeak/IDrop notation for consistency with the companion paper and for conciseness throughout the manuscript, as these indicators are repeatedly used.

<u>R2; C3</u>: "***The dependency on Part 1 is at times inhibiting.The WAI methodology, the foundation for all three indicators, is described only in the broadest strokes here. Readers who have not read Part 1 cannot evaluate whether the indicators are well-defined, whether the wavelet decomposition is appropriate, or whether the edge-padding strategy is valid. In particular:***

- ***The physical meaning of IDrop (ratio of DOG2 April peak to DOG2 November peak at 6-month period) is opaque. I had difficulty understanding what this captures: does a high IDrop value mean a large decline, a steep decline, or something else? Is there always a November peak? A brief schematic or supplementary figure showing the wavelet peaks and how each indicator is extracted would be very helpful.***

- ***The normalization applied for IRise ("scaled between 0 and 1") is mentioned but not explained in detail: normalized across which time period? Does this normalization affect comparability across years?***
- ***The Gaussian window used for IPeak extraction is not described."***

We acknowledge that the reliance on the companion Part 1 paper may limit the accessibility of the present manuscript. While a full description of the WAI framework is beyond the scope of this paper, we have expanded Sect. 2.4 to provide clearer and more self-contained definitions of the three indicators. In particular:

- IRise is now explicitly described as being derived from a yearly normalized GPP signal, ensuring comparability across years.
- IPeak is more clearly defined in relation to the seasonal maximum of GPP, and its computation method using the right lobe of the DOG1 7-month wavelet (which corresponds to a Gaussian window) is better explained.
- IDrop has been reformulated to emphasize its ecological meaning as the relative intensity of the mid-season decline, independent of any secondary peak in the signal.

⇨ Changes in Manuscript: Following the referees' suggestions, we have:

i. Sect. 2.4 (L221–233): clarified definition of all indicators as well as the methodological description of wavelet extraction;
ii. Sect. 2.4 (L234–235): added a reference to Fig. 4 providing a schematic representation of the indicators;
iii. Fig. S1, Sect. 2.4 (L234): added a supplementary figure illustrating the wavelet-based computation of the indicators.

Statistical framework

<u>R1; C2</u>: ***"Statistical methods require some clarification. I don't understand: (i) how longer correlation windows were "decomposed" into shorter windows (L240, L251–254); (ii) how "overlapping correlations" (L243) were defined (for example: what if the window for A overlaps with B and the B window overlaps with C, but C and A don't overlap – are these all in the same group?); or (iii) how the partial correlation analyses were done to identify the most influential variables. What if VPD and Tair both, individually, "rendered remaining [variables] non-significant" (L247) – how would you decide which was more influential?"***

We agree that the original description of the statistical approach was not sufficiently explicit and have substantially revised Sect. 2.5 to improve clarity. The revised description now presents the method as a structured, stepwise procedure, including four main steps:

1. the identification of coherent regions of influence in correlograms across the 2D space of window start dates and lengths, reflecting temporally coherent ecological processes;
2. the decomposition of broad-scale correlations into shorter characteristic windows through the extraction of local maxima;
3. the definition and grouping of overlapping windows, including cases of indirect overlap (e.g. if A overlaps with B and B with C, all belong to the same group);
4. the selection of dominant variables using partial correlations. A variable was considered dominant if it rendered the others non-significant when controlled for. In situations where multiple variables remained significant, they were interpreted as co-dominant, reflecting shared physiological control.

We chose this data-driven, pattern-based approach to prioritize ecological interpretability and parsimony rather than strict predictive optimization.

<u>R2; C1</u>: **"*Multiple comparisons are not addressed. The core analysis computes thousands of Pearson correlation coefficients, roughly one per window length (1-week steps, up to ~25 levels) × start date (1-day steps, ~300+ dates) × number of environmental variables (~6) × three indicators, for both current and previous year. With n = 24, even p ≤ 0.05 thresholds will yield many false positives by chance across this search space. The smoothed correlograms and the requirement for coherent spatial clusters of significance partially mitigate this concern visually, but no formal correction (Bonferroni, FDR, or permutation-based) is described or applied. I suggest that the authors either apply a correction or explicitly justify why the spatial coherence criterion is sufficient, and acknowledge the exploratory nature of the analysis."***

We thank the referee for raising this important point regarding multiple comparisons. We fully agree that the large number of tested correlations increases the risk of false positives if individual p-values are interpreted in isolation. However, our analysis does not rely on selecting isolated significant correlations. Instead, it is based on the identification of spatially coherent patterns in correlograms, which strongly reduces the likelihood of spurious detections. Specifically:

- only clusters of significant correlations that are consistent across contiguous windows (in both time and scale) were considered,
- isolated or non-coherent significant values were systematically discarded,
- identified regions were evaluated based on both effect size (correlation magnitude) and ecophysiological plausibility,
- and all relationships are systematically verified using scatter plots to ensure they are not driven by individual years or outliers.

Because of this workflow, p-values are used primarily as a screening criterion, rather than as a strict inferential threshold. We acknowledge that formal multiple-testing corrections (e.g. Bonferroni or FDR) are commonly recommended. However, in this specific context, such corrections are difficult to apply meaningfully because the tested correlations are highly interdependent due to overlapping time windows and the analysis explicitly aims to detect structured regions of influence. Applying strict corrections would therefore be overly conservative and would eliminate most structured and ecologically meaningful patterns.

⇨ Changes in manuscript: To address these concerns transparently, we have:

i. Sect. 2.5 (L271–307): substantially revised Sect. 2.5 to provide a structured description of the method, clarifying that our approach is pattern-based and exploratory rather than strictly inferential;
ii. Sect. 4.6 (L895–901): explicitly discussed this limitation in the revised manuscript.

Reduced sample size and robustness analysis

<u>R2; C2</u>: **"*Sample size after sequential exclusions becomes very small for some key results. The full dataset is n = 22 usable years. For some analyses the effective n shrinks considerably further: the IPeak* analysis excludes 4 thinning years and 5 extreme-summer years, leaving ~13 years. The headline finding about the P56–63 one-week bud-set bottleneck for IPeak* (r = 0.828, p < 0.001) is based on this reduced dataset. I recommend that the authors state the exact sample size for each analysis and discuss the robustness of the P56–63 result explicitly (e.g., with a leave-one-out sensitivity check).*"**

We thank the referee for highlighting the reduced sample size in subset analyses and for these suggestions. We agree that the exclusion of thinning and extreme years reduces the effective sample size and may affect the robustness of some relationships, including the strong correlation observed for the $P_{56–63}$ window. To address this concern, we performed the requested leave-one-out analysis, in which correlations were recalculated after sequentially removing each year. The relationship remained consistently strong across all iterations, with correlation coefficients ranging from 0.76 to 0.87 and all cases remaining statistically significant ($p < 0.001$). This indicates that the result is not driven by any single year. Given the short duration of this window and the reduced sample size, this analysis provides additional support for the robustness of the identified relationship.

We also note that all relationships were systematically verified using scatter plot inspection to ensure that they were not influenced by outliers or leverage points. In addition, p-values inherently account for sample size through the number of degrees of freedom, providing a complementary indication of statistical support, while acknowledging that this alone may not fully capture robustness in reduced subsets.

⇨ Changes in manuscript: We have:

i. Sect. 3.3.2 (L480–482): added a corresponding statement on leave-one-out robustness analysis in the manuscript;
ii. Tables S1–S5: reported the number of observations used for each correlation to improve transparency;
iii. Sect. 4.6 (L902–904): discussed the reduced statistical power as a limitation.

Connection between context and results

<u>R1; C3</u>: ***"I appreciate your well-researched introduction and discussion, but they often stray quite far from the actual study. Wherever possible, it would be helpful if you can directly connect to your study topic and results. (See line-by-line notes for, e.g., L43, 66, 621–636). In the discussion, I suggest leading with the main result and then following with the (possibly) explanatory details, continuing to reference results as useful."***

<u>R1; L43</u>: **"*Short-term climatic drivers of what?*"**

<u>R1; L66</u>: **"*Re. "how beech responds" – in terms of what? Carbon? Mortality? Something else? This is still very broad.*"**

<u>R1; L621–636</u>: **"*How is this section connected to your analyses?*"**

We agree that some sections of the Introduction and particularly the Discussion could be more explicitly connected to the core results of the study.

Our initial intention was to provide short contextual elements to each subsection in order to place the analyses in a broader ecophysiological context, ensuring accessibility to readers from different communities (e.g. flux science, physiology, ecosystem ecology). However, we acknowledge that this occasionally led to sections that were insufficiently anchored in the results.

⇨ Changes in Manuscript: To improve the clarity of the manuscript while preserving its overall readability and integrative perspective, we have:

i. Sect. 4.1.1 (L548–558); Sect. 4.1.2 (L564–570); Sect. 4.2 (L617–621); Sect. 4.3 (L700–702): revised the manuscript so that each Discussion subsection begins with a clearer reference to the main findings, followed by the corresponding interpretation and contextualization;
ii. Sect. 1 (L43 and L66–67): streamlined suggested statements in the Introduction to improve their direct relevance to the study.

Stand development and methodological limitations

<u>R1; C4</u>: "*Your discussion is thorough in contextualizing results, but it is missing some key points related to the methods. How do you expect that the fact that beech is only ~75 % of the basal area to influence results, if at all? What is the implication, if any, of using the nighttime partitioning method? Why did you choose to test just one year of legacy effect? Could you discuss the variability of timing in the GPP indicators across years and how that might influence the correlations with environmental variables quantified in particular windows?*"

We thank the referee for this insightful comment. We agree that several methodological aspects required further clarification and have addressed them explicitly:

1. Species composition

Beech represents more than 75 % of basal area (> 80 % stems/ha) within the footprint and therefore dominates the GPP signal. Contributions from other species are expected to be limited. The seasonal patterns identified in our analysis are consistent with known beech phenology, supporting this interpretation. Minor influences from co-occurring species (mostly oak and hornbeam) cannot be entirely excluded, particularly for early-season dynamics, but are unlikely to explain the observed magnitude of variations in GPP indicators. In addition, no clear shifts in species-driven canopy dynamics were observed following thinning events, which represent the main driver of structural and compositional changes in the stand, further supporting the limited influence of secondary species on the GPP signal.

2. Partitioning approach

We used the nighttime partitioning method, which is widely applied in eddy covariance studies. We favored it over daytime methods because it minimizes assumptions about GPP response functions and avoids introducing additional model dependencies. While this approach relies on modelling ecosystem respiration and may introduce some uncertainty in absolute flux estimates, its influence on the temporal patterns analyzed here is expected to be limited. This is partly because the indicators used integrate GPP over multi-day to seasonal timescales, which reduces the impact of short-term partitioning errors. Consistently, additional tests using alternative processing approaches (e.g. daytime partitioning or outlier filtering) resulted only in minor shifts in correlation windows and strengths, without altering the overall interpretation.

3. Legacy effects

We tested legacy effects over one year only, assuming that the strongest impacts occur closest in time to the stress event. Extending the analysis to multi-year lags would reduce statistical power given the limited sample size. This choice is now clarified in the manuscript.

4. Timing variability

GPP indicators are derived using a wavelet-based approach that aligns them with intrinsic discontinuities in the GPP signal, allowing adaptation to interannual phenological shifts. Environmental drivers are also computed relative to the start of season (SOS), ensuring temporal alignment between drivers and ecosystem response. This approach implicitly assumes that phenological timing relative to SOS is broadly consistent across years. This hypothesis is now explicitly stated in the manuscript.

<u>R2; C4</u>: "*Forest maturation versus climate change in the trend analyses. The forest was approximately 31 years old at the start of the record (1997) and 55 years old at the end (2020). This is a period of*"

substantial stand development: LAI, canopy height, and stand structure all evolved considerably, and thinning events further altered stand composition. The paper discusses thinning effects but does not explicitly address whether the significant positive trend in IPeak and the negative trend in IRise could partly reflect forest ontogeny (increasing canopy closure, shift from rapid leaf expansion to sustained peak production as the canopy matures) rather than climate forcing alone. The extreme IPeak years (2013–2020) cluster at the end of the record, when the forest is also oldest. Could the authors test the effect of removing this trend?"

We thank the referee for this insightful comment. We fully agree that forest maturation may influence long-term trends in GPP indicators and that this aspect required further clarification.

To address this point, we have explicitly assessed the role of interannual variability by recalculating temporal trends after excluding extreme summer years. In this case, trends in both IPeak and IRise became non-significant (IPeak: $r = 0.37$, $p = 0.14$; IRise: $r = 0.3$, $p = 0.25$), indicating that the apparent long-term patterns are largely driven by extreme events rather than gradual stand development. This interpretation is supported by additional observations: (1) leaf area index did not show a clear increasing trend over the study period, suggesting that canopy closure was largely achieved early in the record (approx. by 2001); and (2) biomass increment declined after 2013, indicating that the stand had reached a mature phase rather than undergoing continuous structural development.

We therefore interpret the observed temporal variability primarily as climate-driven, while acknowledging that structural and ontogenetic processes may contribute secondarily.

<u>R2; C5</u>: "*The negative IRise trend could have more discussion. The significant decline in IRise over 24 years (Fig. 3) is noted but its ecological interpretation is discussed leak than IPeak. As the stand matures and canopy closure increases, one might expect slower early-season GPP rise (less light penetration to understory, more self-shading) alongside higher peak productivity. This maturation interpretation could be contrasted against alternatives (e.g., earlier SOS reducing the "room to rise"). The negative IRise/IPeak correlation (stronger without thinning years) suggests these two indicators trade off against each other in a way that may be structurally driven (relating to forest maturation) rather than climatically driven."*

We thank the referee for this valuable suggestion. We agree that the ecological interpretation of the negative IRise trend required further development.

Our additional analyses indicate that this trend is largely influenced by the combined effects of earlier SOS (detailed in the companion paper) and interannual variability in peak productivity (see Fig. S4). An earlier onset of the growing season increases the time available to reach peak GPP, effectively reducing the apparent rate of increase (IRise), a "room-to-rise" effect. This interpretation is consistent with the advancement of SOS, further discussed in Part 1. The negative relationship between IRise and IPeak further suggests an interaction between early-season dynamics and peak productivity, potentially reflecting both climatic variability and internal ecosystem regulation. While we acknowledge that forest maturation may contribute (e.g. through changes in canopy structure), our results indicate that these effects are secondary compared to climatic variability, particularly the influence of extreme years.

⇨ Changes in manuscript: To address these points more explicitly, we have:

i. Sect. 4.5.2 (L877–894): added a dedicated subsection discussing structural changes and forest maturation (including the IRise–IPeak relationship);
ii. Sect. 4.6 (L905–919): added a dedicated limitations section addressing these aspects (four points of **R1; C4** and climatic/ontogenic effects for **R2; C4**).

Minor changes

R1; L18: ***"At this stage in the manuscript, it's not clear why $VPD_{92–106}$ and $VPD_{107–114}$ would be considered separate windows."***

⇨ Changes in Manuscript (L18–19): We have modified the sentence in the Abstract to distinguish the two windows as reflecting pre-drop and post-drop onset phases of stress.

R1; Figure 1: ***"Nice figure! Please connect the references with the processes using, for example, superscripts."***

Thank you for your positive comment and for this suggestion.

⇨ Changes in Manuscript (Fig. 1; L73–77): We have added connections between references and processes in Fig. 1.

R1; L110: ***"What are "Warm Winter records"?"***

"Warm Winter records" refers to a specific dataset produced for the ICOS processing chain.

⇨ Changes in Manuscript (L112): We simplified this sentence to avoid unnecessary terminology and now refer only to "ICOS-processed data" which keeps the description concise.

R1; L128–133: ***"Provide equations please (and number all equations, to be referenced in the text)."***

⇨ Changes in Manuscript (L128–138): We have added Eqs. (1)-(4), illustrating index computation for a non-bissextile year y, and numbered all equations consistently throughout the manuscript.

R1; L151–154: ***"Provide equations please"***

⇨ Changes in Manuscript (L160–166): We have added the corresponding equations, Eqs (6)-(7), and clarified the definitions of all derived variables.

__**R1; L175–177**__: **“*What sensors were used to measure albedo and NDVI?*”**

Albedo was evaluated as the ratio of reflected to incoming short-wave radiation, both measured using a pyranometer. NDVI was computed as the difference between near-infrared and red reflectance divided by their sum. Reflectance at each wavelength was measured using sensors recording incoming and reflected radiation in the red (approx. 650 nm) and near-infrared (approx. 810 nm) spectral bands.

⇨ Changes in Manuscript (L189–192): We have clarified the description of albedo and NDVI by explicitly specifying how the relevant radiation components were measured and derived at the canopy level.

__**R1; L182–185**__: **“*Why were there times when the GPP definition wasn’t available? Did these definitions, when multiple were available, typically agree (this could be in a supplement)?*”**

We clarify that GPP itself was always available from flux partitioning; the reported percentage refers to the availability of independent SOS detection metrics (RLT, albedo, NDVI).

These metrics were not used as independent SOS estimates but as sequential consistency filters to identify a single date combining canopy structural change (RLT) and spectral response (NDVI/albedo).

Because the objective was not to compare methods but to ensure a robust and ecophysiologically consistent SOS detection, we did not perform a systematic comparison between metrics.

⇨ Changes in Manuscript (L195–198): We have clarified these points in Sect. 2.3.3 to avoid ambiguity.

__**R2; L195–200**__: **“*The WAI description mentions the "companion paper that forms the first part of this study". I’d provide the full citation here so readers can locate it.*”**

⇨ Changes in manuscript (L212): We have added the full citation of the companion paper (preprint) to facilitate access.

__**R1; L220**__: **“*Are all the correlations linear?*”**

Yes, all correlations are linear (Pearson). We chose Pearson correlations because they provide direct interpretability, are consistent with the use of partial correlations, and yielded results very similar to rank-based approaches (Spearman) in this dataset.

⇨ Changes in Manuscript (L247–248): We have clarified the definition of both methods in Sect. 2.5.

R1; L229: ***"I suggest a "for example" parenthetical here to clarify: "This approach produced one average value per year for each window configuration (for example, one VPD value for a week-long window beginning at SOS+0, one for a week long window beginning SOS+1, etc.).""***

Thank you for this suggestion.

⇨ Changes in Manuscript (L257–258): We have added the suggested example.

R1; L256: ***"Residuals of what model?"***

R1; L306: ***"Again, I'm confused by where these residuals are coming from."***

The residuals are derived from simple linear regressions between each indicator and its associated environmental driver (i.e., the same relationships used in the Pearson correlation analysis). Residuals therefore represent deviations from the expected linear response of the ecosystem to environmental conditions.

⇨ Changes in Manuscript (L308–309): We have clarified this point in Sect. 2.5 to avoid confusion.

R2; L286: ***"The section 3.1.1. & 3.1.2 headings are duplicated between pages 10 & 11"***

⇨ Changes in manuscript (L338): The section title has been corrected to "3.1.2 Temporal trends and relationships among GPP indicators" to accurately reflect its content.

R2; L287: ***""IPeak exhibited a strong positive trend". I'd add the slope and p-value explicitly in the text, not just in Figure 3."***

R2; L288: ***""IDrop showed no significant long-term trend". I'd include the actual p-value here."***

⇨ Changes in manuscript (L339–341): We have added the slope and p-value for IRise, IPeak and IDrop directly in the text.

__R1; Figure 3__: **“*What are the units for IRise, IPeak, and IDrop? The second column of figures is NOT residuals from the models in the first column, which is confusing… where do these residuals come from? Consider labeling all the years, not just extreme years. Re. "Years affected by thinning" – what makes you think that the thinning is only influential in the year it is done?*”**

Thank you for these helpful remarks.

- IRise, IPeak and IDrop are derived from wavelet coefficients. Because the wavelet transform preserves the dimensionality of the original signal (using a L1-normalisation), IPeak is expressed in the same units as GPP (gC m^{-2}), while IRise (computed from a normalized signal) and IDrop (defined as a ratio of wavelet coefficients) are dimensionless.
- We clarified that the residuals are derived from separate linear relationships, not from the panels shown alongside them.
- While labeling all years would improve completeness, it resulted in substantial visual overcrowding and reduced readability of the figure. We therefore retained labels for extreme years only, as they are central to the interpretation of the results.
- We also rephrased “years affected by thinning” to “years during which thinning operations were conducted,” to avoid implying that effects are limited to that year.

⇨ Changes in Manuscript: We have:

i. Fig. 3 (L347 and L350–353): improved the caption accordingly;
ii. Fig. 3, Fig. 6, Fig. S4, Sect. 2.4 (L235–237): added units in all figures, while stating in the captions and in Sect 2.4 that indicators should be interpreted in a relative sense (e.g., across years or events).

__R2; L300–303__: **“*The negative IRise–IPeak correlation is discussed as possibly reflecting "earlier saturation or stress onset." I would be good to mention other explanations here, too (canopy structure, leaf area development trade-offs, forest maturation).*”**

⇨ Changes in manuscript (L357–358): We have expanded the discussion of the negative IRise–IPeak relationship to include additional potential explanations.

<u>R1; Figure 4</u>: "*What a great figure! It really gets a lot of information across in an intuitive way. Would it be useful to scale the thickness of the bars to the correlation coefficient (maybe not?). How did you decide how to order the correlation bars from top to bottom within each section? Consider ordering by correlation magnitude or by beginning date of the correlation window. **Which correlations are shown? Sometimes, in the text, windows are referenced as though they're significant but they're not in this figure – e.g. L355, L382, L427–430, L518. Or, windows are in the figure but are omitted from the text as though they're not significant – e.g. L413.*"

Thank you for this comment and for the positive feedback on the figure.

Figure 4 displays only the dominant correlations selected through our variable-selection procedure (see Methods), which explains why some significant windows mentioned in the text are not shown.

Regarding the specific examples:

- L355, L413, L518: the windows $P_{95–137}$, $P_{57–120}$, $Tair_{22–43}$ and $Tair_{9–16}$ are included in Fig. 4.
- L427–430: only $VPD_{53–186}$ was retained after variable selection; we also corrected an inversion in the starting dates for P (SOS+58) and VPD (SOS+53) in L429–430.
- L382: this window was not retained because it could be explained by other variables; we have clarified this explicitly in the text.

In the figure, long-scale windows are systematically displayed above short-scale ones, which are displayed chronologically (except a minor ordering error regarding the $DL_{9–16}$ window for IRise). We will try implementing suggested adjustments to improve visual readability (in particular, correlation strength).

⇨ Changes in Manuscript: We have:

i. Fig. 4: clarified the ordering of correlations in the caption and corrected the ordering of short-scale windows for IRise;
ii. Sect. 3.3.1 (L440): added a clarification regarding the rejection of the $Rg_{151–158}$ window.
iii. Sect. 3.4.1 (L494–497): corrected starting dates for P and VPD;

<u>R1; L379</u>: "*The Rg significant window was actually 9 days later than maximum incident radiation; does this really correspond?*"

We agree that the term "correspond to" was not appropriate.

⇨ Changes in Manuscript (L435): We replaced it with "are linked to," to better reflect the temporal offset while maintaining the ecological interpretation.

<u>R1; L385–386 (Figure 6)</u>: "*I don't see any thresholds in these figures…*"

<u>R1; L387</u>: "*Consider labeling 2003 on Fig. 6*"

Threshold values were not reported in the initial version because they were not explicitly quantified through a formal threshold analysis. However, as they help interpret the observed relationships, we now provide approximate values (250 W m^{-2} for Rg, 16 °C for Tair, and 10 hPa for VPD) and illustrate both full and filtered relationships (i.e. excluding summer-extreme years) within the same figure.

⇨ Changes in Manuscript: We have:

i. Fig. 6: added filtered relationships and specific labelling for 2003;
ii. Fig. 6; Sect. 3.3.1 (L442–444): provided approximate threshold values directly in the figure caption and corresponding section of the manuscript.

<u>R1; Figure 7</u>: "*This is a particularly information-light figure; add all year labels; consider scatter plots with Reco, g_c on the y axis and extremity on the x axis.*"

Thank you for this suggestion. We updated Figure 7 to improve its informativeness and readability, including labeling all years directly and clarifying the ranking of variables. During this revision, we identified an inconsistency in the original version: bars associated with g_c were inadvertently derived from latent heat data rather than g_c. This has been corrected. The overall interpretation of the results remains unchanged.

⇨ Changes in Manuscript: We have:

i. Fig. 7: updated the figure accordingly;
ii. Sect. 3.3.1 (L467–468); Sect. 4.2 (L622–623): revised the corresponding statements in the manuscript to reflect the updated pattern, emphasizing that canopy conductance is generally high across years but particularly elevated in 2013, 2014 and 2018, rather than consistently across all extreme years.

<u>R1; L424</u>: "*But none of the extreme years were next to each other except 2018/19*"

We agree with this observation.

⇨ Changes in Manuscript (L489–490): We removed the sentence for clarity as it did not add substantial value to the interpretation.

<u>R1; L426</u>: "*I don't see "Rg began correlating with IDrop from SOS+8 onward" in either Fig. 4. Or Fig. S3 – what am I missing?*"

Thank you for pointing this out.

⇨ Changes in Manuscript: We have:

i. Sect. 3.4.1 (L492): referred to the appropriate figure (Fig. S9) and table (Table S5);
ii. Sect. 3.4.1 (L501 and L505): clarified references to figures and tables for IDrop.

<u>R1; L446–449</u>: "*Why should stress AFTER the drop influence the drop?*"

We clarify that this second window refers to conditions occurring after the onset of the drop, not after its completion. These post-onset conditions influence the intensity and duration of the decline, rather than triggering it. We further discuss the relevance of this window in L662–663.

⇨ Changes in Manuscript (L510): We have revised the text accordingly.

<u>R1; L455</u>: "*This seems to conflict with L438.*"

We acknowledge that the wording was unclear. VPD is the dominant driver over most of the period (see correlation values, number of selected VPD windows in Fig. 4 / Table S5 and response to L433), while REW plays a secondary and more localized role, particularly late in the season.

⇨ Changes in Manuscript (L520–526): We have revised this paragraph to clarify this distinction.

<u>R1; L513–514</u>: "*I don't really understand this.*"

⇨ Changes in Manuscript (L589–591): We have reformulated this statement to explain that environmental stress can slow leaf expansion rates while simultaneously extending the duration of expansion (as described in Granier and Tardieu, 2009). As a result, final leaf area may remain unchanged despite reduced instantaneous growth rates.

R1; L541: **"*It looks like pretty strong coupling in Figure 6 to me.*"**

R1; L542: **"*2013, 2014, 2018, 2019, 2020 are not notably less scattered around the fit line than the other years.*"**

We clarify that the "weak coupling" statement refers to typical years, when excluding extreme-summer years.

⇨ Changes in Manuscript (L617–621): We now explicitly refer to the relationships observed when excluding extreme years to avoid confusion and rely on the updated Fig. 6 to better illustrate the linear increase.

R1; L644: **"*Higher sensitivity to atmospheric demand than to what?*"**

⇨ Changes in Manuscript (L724–725): We have clarified the sentence by specifying that beech at mesic sites shows higher sensitivity to atmospheric demand than to soil water limitation (which is consistent with recent literature, e.g. Green, 2024; Novick et al., 2016).

R1; L645–646: **"*This is a convoluted sentence, and I don't think it's supported by the fact that VPD correlates well with IDrop.*"**

We agree that the original formulation was too categorical.

⇨ Changes in Manuscript (L725–727): We have revised the sentence to better reflect that the relative importance of soil versus plant hydraulic limitation is context-dependent. The revised wording emphasizes that our interpretation applies specifically to site conditions and is consistent with theoretical frameworks describing how the limiting hydraulic element governs sensitivity to VPD versus soil drying (Wankmüller et al., 2024).

R1; L779–784: **"*I don't think this all follows. Severe thinning reduces IPeak whereas minor thinning doesn't; how does this suggest that loss of photosynthetic surface outweighs effects of increased light availability and competition release? – this would only be the case for severe thinning, presumably.*"**

We agree that the original formulation was unclear and could be misinterpreted. We have revised the paragraph to clarify that the observed pattern reflects a contrast between thinning intensities. Severe thinning (e.g., 2005) leads to a strong reduction in photosynthetic surface that outweighs potential gains in light availability and reduced competition. In contrast, lighter thinning events induce only minor structural changes, which are insufficient to significantly affect stand-level productivity.

⇨ Changes in Manuscript (L857–860): This distinction is now made explicit in Sect. 4.5.1.

R1; TableS1: “***What does bold mean? It’s not clear which effects are dominant.***”

⇨ Changes in Manuscript: We have:

i. Tables S1-S5: replaced the use of “bold” formatting with “blue” to improve clarity;
ii. Table S1: updated the caption accordingly;
iii. Table S5: corrected minor inconsistencies (grouping, correlation values and windows), which do not affect the interpretation of the results.

R1; L63: “***influences → influence***”

R1; L127: “***aerial drought → atmospheric drought***”

R1; L244: “***influent → influential***”

R1; L303: “***indicating → suggesting***”

R1; L461: “***show with → have***”

R2; L659: “***"a second, shot period": "short"?***”

R1; L660: “***shot → short***”

R1; L664: “***aerial → atmospheric***”

R1; L671: “***drought → droughts***”

⇨ Changes in Manuscript: Corrected.

Responses without changes

R1; L342: "***Rg**$_{15–29}$ **is also in Fig. 4 (in addition to Rg10–31), and it has a larger correlation than Rg**$_{10–31}$ **– why isn't is considered here?***"

The Rg$_{15–29}$ window is indeed shown in Fig. 4, although it has a lower correlation with IRise (0.741) than Rg$_{10–31}$ (0.782). Both windows were considered; however, Rg$_{15–29}$ represents a shorter sub-period nested within the broader Rg$_{10–31}$ window and was therefore only discussed in relation to its second phase (SOS+15 to SOS+30). The broader Rg$_{10–31}$ window, which integrates both phases, is discussed in L334–340.

R1; L413: "***What about prior year P**$_{56–63}$**?***"

This effect is discussed in the manuscript but embedded within the broader P$_{57–120}$ window (see L414).

R1; L433: "***But VPD and REW are correlated (and causally related) – how did you decide which of these was more important (see broad note about statistical method clarification)***"

The distinction between VPD and REW was based on a combination of (1) correlation strength and (2) partial correlation analysis, which allowed us to assess whether one variable explains the signal of the others (see response to broad note).

In this case, VPD consistently showed stronger correlations and remained significant when controlling for REW, whereas the reverse was not true (see table below). This indicates that VPD is the dominant driver of IDrop, while REW plays a secondary role.

Control: VPD				**Control: REW**			
Variable	Partial correlation	p-value	Significance level	**Variable**	Partial correlation	p-value	Significance level
Rg	−0.06	0.80	/	**Rg**	0.54	0.016	*
P	−0.44	0.06	/	**P**	−0.37	0.12	/
Tair	0.11	0.66	/	**Tair**	0.61	0.006	**
REW	−0.27	0.26	/	**VPD**	0.76	0.0002	***
