Author comment: response to referee n°2

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)

We sincerely thank the referees for their thorough and constructive evaluation of our manuscript. Their comments have been extremely helpful in improving the clarity, structure, and overall quality of the work.

This document provides a detailed, point-by-point response to all comments. We have addressed each referee separately, while ensuring consistency in the revisions throughout the manuscript. For transparency, responses are organized as follows: an overall assessment, followed by general comments and, where applicable, specific line-by-line comments as well as comments related to figures and tables.

Although some comments addressed similar aspects of the study (e.g., indicator definitions and statistical approach), we retained individual responses for each referee, as their remarks were complementary and led to specific revisions in different parts 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.

Overall assessment

"***This is an ambitious and successful paper. The authors bring 24 years of high-quality eddy covariance data to bear on a genuinely important question: how do brief, seasonally specific environmental windows shape annual forest carbon uptake? The phenology-aligned sliding window correlation framework is creative, the discussion is impressively well-grounded in beech physiology, and Figure 4 in particular is a model of information-dense, intuitive visualization. I expect this will be a well-cited contribution to forest carbon cycling literature. Below I detail a handful of suggestions and questions regarding the statistical approach and methodological transparency, which I believe the authors can address in a minor revision. I provide these suggestions with the hope that they help clarify sections of the paper I found more opaque, and make the contribution stronger in its final published form.***

Overall, this is a well-conceived study with real scientific value. The findings are interesting and I thought that the figures were very well made. I believe the work should be published in Biogeosciences, and I hope that the suggestions that I've provided here help strengthen the final contribution. I look forward to seeing this work in print soon."

We thank the referee for the positive and thoughtful assessment of our work. We particularly appreciate the recognition of the methodological approach and the ecological interpretation of the results. The referee's comments helped us improve the transparency and robustness of the study, especially regarding (i) the treatment of multiple comparisons and the robustness of key relationships, (ii) the role of forest structural effects and its interaction with climatic drivers and (iii) the accessibility of the wavelet-based methodology. We believe that these revisions have strengthened both the methodological rigor and the ecological interpretation of the manuscript.

General comments

<u>Comment n°1</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 this concern transparently, we have:

i. clarified in Sect. 2.5 that our approach is pattern-based and exploratory rather than strictly inferential;
ii. explicitly discussed this limitation in the revised manuscript (Sect. 4.6).

<u>Comment n°2</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. added a corresponding statement on leave-one-out robustness analysis in the manuscript (Sect. 3.3.2);
ii. reported the number of observations used for each correlation in Tables S1–S5 to improve transparency;
iii. discussed the reduced statistical power as a limitation (Sect. 4.6).

<u>**Comment n°3**</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: We have:

i. expanded accordingly definitions of IRise, IPeak and IDrop and clarified the methodological description of wavelet extraction (keeping the level of detail appropriate for the manuscript) in Sect. 2.4;
ii. added explicit reference to Fig. 4 as a schematic illustration of how the indicators relate to the seasonal GPP trajectory, improving readability for readers unfamiliar with the WAI framework;
iii. added a supplementary figure illustrating the wavelet positioning for computing indicators (see Fig. S1 below).

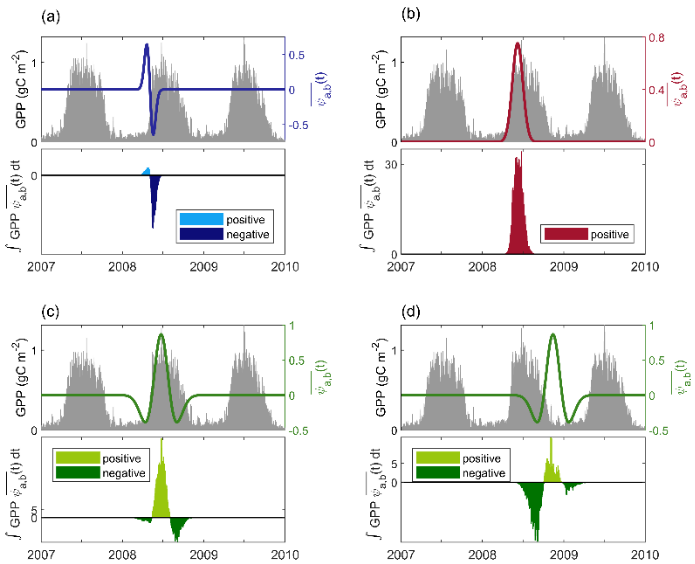

Figure S1: Computation steps for deriving (a) IRise, expressed as the DOG1 April peak at a 3-month period on a normalized GPP signal, (b) IPeak, computed using the opposite of the right lobe of the DOG1 April peak, which is equivalent to a Gaussian function, at a 7-month period, and (c)/(d) IDrop, defined as the ratio of the DOG2 (c) June to (d) November peaks at 6-months periods. For each panel, the top graph displays the wavelet at selected scale and time with the GPP signal in the background (restricted to positive values for readability). The graph below illustrates the area under the product curve of the wavelet and the signal. The algebraic sum of the contributing areas, each linked to wavelet lobes, corresponds to the indicator value yielded by the WT (after taking the ration c/d for IDrop). Computations are presented for the year 2008.

<u>Comment n°4</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.

⇨ Changes in manuscript: To address this point more explicitly, we have:

i. added a dedicated subsection (Sect. 4.5.2) discussing structural changes and forest maturation;
ii. clarified that disentangling climatic and ontogenetic effects remains a limitation of long-term observational studies (Sect 4.6).

<u>Comment n°5</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. S2 – Fig. S4 in the revised manuscript). 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: These interpretations have been incorporated into the revised manuscript, within the new Sect. 4.5.2.

Line-by-line comments

<u>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. 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;
ii. added a dedicated Supplementary section (Sect. S2) and figure (Fig. S2) describing the equations and their validation in detail.

<u>L195–200</u>: **“*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: We have added the full citation of the companion paper (preprint) to facilitate access.

<u>L286</u>: **“*The section 3.1.1. & 3.1.2 headings are duplicated between pages 10 & 11*”**

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

<u>L287</u>: **“*"IPeak exhibited a strong positive trend". I’d add the slope and p-value explicitly in the text, not just in Figure 3.*”**

⇨ Changes in manuscript: We have added the slope and p-value for IPeak directly in the text. For consistency, we also included the corresponding values for IRise.

<u>L288</u>: "*"IDrop showed no significant long-term trend". I'd include the actual p-value here.*"

⇨ Changes in manuscript: We have added the slope and p-value for IDrop directly in the text.

<u>L300–303</u>: "*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: We have expanded the discussion of the negative IRise–IPeak relationship to include additional potential explanations both in L303 and in Sect. 2.5.2 (see responses to general comments).

<u>L659</u>: "*"a second, shot period": "short"?*"

⇨ Changes in manuscript: Corrected.
